

数学的源与流

（第二版）

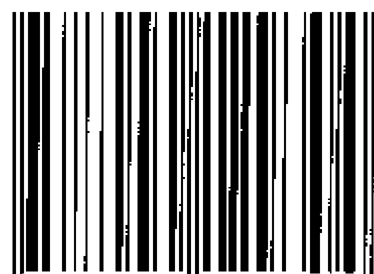
张顺燕　编著

高等教育出版社

SHUXUEDEYUANYULIU

责任编辑：胡乃冏
封面设计：于　涛

ISBN 7-04-012930-2



定价 23.90 元

数学的源与流

第二版

张顺燕　编著

高　等　教　育　出　版　社

图书在版编目（CIP）数据

数学的源与流/张顺燕编著.—2版.—北京:高等教育出版社,2003.12

ISBN 7-04-012930-2

Ⅰ.数… Ⅱ.张… Ⅲ.数学-高等学校-教材
Ⅳ.01

中国版本图书馆CIP数据核字(2003)第089361号

出版发行	高等教育出版社	购书热线	010-64054588
社　　址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010-82028899		http://www.hep.com.cn

经　　销	新华书店北京发行所
印　　刷	北京二二〇七工厂

		版　　次	2000年9月第1版
开　　本	850×1168　1/32		2003年12月第2版
印　　张	17.75	印　　次	2003年12月第1次印刷
字　　数	220 000	定　　价	23.90元

内容介绍

本书是北京大学数学素质教育课的主要教材.

内容包括著名的数学问题、具有重要实用价值的应用问题，还包括数学的一些近代应用.

本书立意新颖、内容丰富、涵盖面广、观点高、起点低，只要具备中等数学的基础就能读懂大部分内容；最后几章要用到初等微积分.本书可作为大专院校数学素质教育的参考书，对广大中学数学教师提高数学素养也极有参考价值.

序

《数学的源与流》是张顺燕教授经过长期思考写的一本向普通读者介绍数学的科普读物.数学处理的对象是抽象的,举一个简单的例子,初等几何研究的“直线”就是一个抽象的概念,它与实际存在的直线不同,没有宽度.因为数学处理的对象是抽象的,在数学的发展中常常是一个定义接一个定义,而抽象程度又一层层地提高,所以数学很难为一般人所理解.随着科学技术对人类生产和生活产生日益深刻的影响,数学的重要性越来越清楚地显示出来,而且数学应用的领域也越来越广泛,但数学为什么会有用,以及数学在人的教育方面的作用究竟有多大,却是不容易说清楚的,要全面的介绍数学究竟是什么,在目前也几乎是不可能的.张顺燕教授为了使普通读者对数学有一个比较完整的了解,选择了梳理数学的“源”与“流”这样一种角度,“源”讲数学的发生,数学的一些概念开始是如何从客观实际抽象出来的;“流”讲数学的发展,讲数学在2000多年的时间里,伴随着人类知识的不断积累和思维能力的提高,数学概念层层抽象,数学研究的范围不断扩大,研究对象不断变化的过程。我认为本书选取这样一个角度是颇具慧心的,书中对介绍数学知识的各个部分的选择也比较恰当,对于一些与实际联系紧密,比较容易叙述清楚的东西,就说得详细一点,对于一些特别抽象,难于说清楚的东西,就介绍得简略一些,这样就避免了在复杂的问题上纠缠不清的缺点.张顺燕教授也考虑到数学的现代发展,在书中也涉及到了数学的一些新的分支,比如数学界最近讨论比较多的混沌理论.总之,这本书有简有繁,有深有浅,力图使读者对数学有一个总体轮廓的了解。我想这个目标是非常有意义的,也是很难达到的,张顺燕教授在这方面进行了比较成功的尝试.希望本书的出版,对普及数学知识,对提高大家对数学重

要性的认识能够发挥积极的作用.记得有位著名数学家说过,今天数学教育水平的高低,就决定了明天科学技术的发展.这句话被历史证明是正确的.我衷心希望大家能够更加重视数学教育,希望中国数学教育水平在新世纪有大的提高.

丁石孙

2000.12.26

前　言

1998,1999两年,笔者在中央教育电视台录制了"今日数学"讲座.对象是广大的中学数学教师.目的有二:

1.介绍数学思想的演变和发展,加深对数学思想的理解;

2.了解重大数学成果,及数学对人类文明的贡献,以达到开拓视野,启发灵性的效果.

现在的书名是高等教育出版社的张小萍等先生拟定的,比原名更好.但是,单靠书名来概括全书,这是难以做到的.还是要请读者看具体内容.

目前素质教育正在全国兴起,今年秋季北京大学将向全校(不分文、理,不分年级)开设数学素质教育课.数学学院领导建议由笔者来开这门课.笔者欣然接受,愿为素质教育效一份力.此书就成为这个课的最主要的教材,课程的大部分内容将从中选取.

本书也可供兄弟院校作为数学素质教育的教材,并根据本校的具体情况选取所需内容.

高教出版社的胡乃冋先生对此给予以大力支持.作为急件,加速排版,力争在开学初出书,表现了出版界对素质教育之重视.

在讲座拍摄过程中,姜伯驹院士、武际可教授、李忠教授、潘承彪教授都曾给予鼓励与支持.姜伯驹院士建议讲《一笔画与邮递路线问题》,并慨然将他的名著提供给笔者.他的著作简单、精要,笔者基本上按他的著作讲授,只是后面作了一些引申.

全国人大副委员长丁石孙教授对讲座甚为关怀,尤为使笔者感动的是,他的藏书供笔者自由选用.1999年春节一次就从他家借书10本,用了一年之久,使笔者收益甚大.他主编的《数学、我们、数学》丛书即在其中.这套丛书论述了数学与社会各个领域的联系,价值很

高，开了素质教育之先河.

本书需要的预备知识不多，大部分内容只要有高中数学的知识就可读懂.最后几章需要初等微积分的知识.

限于笔者的水平，错误是一定有的，欢迎读者批评指正.

张顺燕 于北京大学燕北园

2000 年 8 月 18 日

策划编辑　王　瑜

责任编辑　胡乃冏

封面设计　于　涛

责任绘图　朱　静

责任印制　宋克学

目　　录

第一章　数学与人类文明

数学是科学的大门和钥匙.

Rogen Bacon

所以这就是数学:它赋予自己的发现以生命;它令思维活跃,精神升华;它烛照我们的内心,消除了我们与生俱有的蒙昧与无知.

Proclus

古今之成大事业大学问者,必经过三种之境界."昨夜西风凋碧树,独上高楼,望尽天涯路",此第一境界也."衣带渐宽终不悔,为伊消得人憔悴",此第二境界也."众里寻她千百度,蓦然回首,那人却在灯火阑珊处",此第三境界也.

王国维

1.1.1　数学的内容

大致说来,数学分为初等数学与高等数学两大部分.

初等数学中主要包含两部分:几何学与代数学.几何学是研究空间形式的学科,而代数学则是研究数量关系的学科.

初等数学基本上是常量的数学.

高等数学含有非常丰富的内容,以大学本科所学为限,它主要包含:

解析几何:用代数方法研究几何,其中平面解析几何部分内容已放到中学.

线性代数:研究如何解线性方程组及有关的问题.

高等代数:研究方程式的求根问题.

微积分:研究变速运动及曲边形的求积问题.作为微积分的延伸,物理类各系还要讲授常微分方程与偏微分方程.

概率论与数理统计:研究随机现象,依据数据进行推理.

所有这些学科构成高等数学的基础部分,在此基础上建立了高等数学的宏伟大厦.

1.1.2 数学的特点

数学区分于其它学科的明显特点有三个:第一是它的抽象性,第二是它的精确性,第三是它的应用的极端广泛性.

从中学数学的学习过程中读者已经体会到数学的抽象性了.数本身就是一个抽象概念,几何中的直线也是一个抽象概念,全部数学的概念都具有这一特征.整数的概念,几何图形的概念都属于最原始的数学概念.在原始概念的基础上又形成有理数、无理数、复数、函数、微分、积分、n 维空间以至无穷维空间这样一些抽象程度更高的概念.但是需要指出,所有这些抽象度更高的概念,都有非常现实的背景.不过,抽象不是数学独有的特性,任何一门科学都具有这一特性.因此,单是数学概念的抽象性还不足以说尽数学抽象的特点.数学抽象的特点在于:第一,在数学的抽象中只保留量的关系和空间形式而舍弃了其它一切;第二,数学的抽象是一级一级逐步提高的,它们所达到的抽象程度大大超过了其它学科中的一般抽象;第三,数学本身几乎完全周旋于抽象概念和它们的相互关系的圈子之中.如果自然科学家为了证明自己的论断常常求助于实验,那么数学家证明定理只需用推理和计算.这就是说,不仅数学的概念是抽象的、思辨的,而且数学的方法也是抽象的、思辨的.

数学的精确性表现在数学定义的准确性、推理的逻辑严格性和数学结论的确定无疑与无可争辩性.这点读者从中学数学就已很好地懂得了.当然,数学的严格性不是绝对的,一成不变的,而是相对的,发展着的,这正体现了人类认识逐渐深化的过程.

数学应用的极其广泛性也是它的特点之一.正像已故著名数学家华罗庚教授曾指出的,宇宙之大,粒子之微,火箭之速,化工之巧,地球之变,生物之谜,日用之繁,数学无处不在,凡是出现“量”的地

方就少不了用数学,研究量的关系,量的变化,量的变化关系,量的关系的变化等现象都少不了数学.数学之为用贯穿到一切科学部门的深处,而成为它们的得力助手与工具,缺少了它就不能准确地刻画出客观事物的变化,更不能由已知数据推出其它数据,因而就减少了科学预见的可能性,或减弱了科学预见的精确度.

1.1.3　数学对人类文明的贡献

数学是什么,它对人类文明有怎样的贡献?这是一个带有根本性质的问题,是一个从事数学教育与数学研究的人应当回答的问题.人类认识的发展基于经验的积累和理性的思维.单靠经验的积累,有时像在黑暗中摸索,不可能有认识上的重大突破.在经验积累的基础上,经过理性的思维才能产生伟大的飞跃.下面举几个对人类文明具有重大影响的例子.

万有引力定律.基于开普勒行星运动的三大定律,牛顿发现了万有引力定律.这是人类对宇宙认识的一次伟大革命.牛顿把他最重要的著作命名为"自然哲学的数学原理",是因为他发现新宇宙的思维方式是数学的思维方式.

相对论.爱因斯坦的相对论是宇宙观的另一次伟大革命,其核心内容是时空观的改变.牛顿力学的时空观认为时间与空间不相干,伽利略变换式是这种数学模型的基本表现形式.爱因斯坦的时空观却认为时间和空间是相互联系的.四维空间的洛仑兹变换是这种数学模型的表现形式.促使爱因斯坦作出这一伟大贡献的仍是数学的思维方式.

海王星的发现.这个太阳系的最远的行星之一,是在1846年在数学计算的基础上发现的.天文学家阿达姆斯和勒未累分析了天王星的运动的不规律性,得出结论:这种不规律性是由其他行星的引力而发生的.勒未累根据力学法则和引力法则计算出这个行星应该位于何处,他把这个结果告诉了观察员,而观察员果然在望远镜中在勒未累指出的位置看到了这颗行星.这个发现是数学计算的胜利.

另一个著名的例子是电磁波的发现.英国物理学家麦克斯韦概括了由实验建立起来的电磁现象规律,把这些规律表述为"方程的形式",他用纯粹数学的方法从这些方程推导出可能存在着电磁波并且这些电磁波应该以光速传播着.据此,他提出了光的电磁理论,这个理论后来被全面地发展和论证了.除此之外,麦克斯韦的结论还推动了人们去寻找纯电起源的电磁波.例如,由振动发电所发射的电磁波.这样的电磁波果然为赫兹所发现.而不久之后,波波夫就找到了电磁振荡的激发、发送和接收的办法,并把这些办法带到许多应用部门,为无线电技术奠定下基础.现在已进入信息时代,无线电技术对于人类生活是何等重要人人都已体会到.但是我们可不要忘记,纯粹数学在这里曾起过巨大作用.

另一个光辉的例子是非欧几何的诞生.它是从欧几里得时代起的几千年来从人们想要证明平行公设的企图中,也就是说,从一个只有纯粹数学趣味的问题中产生的.罗巴切夫斯基创立了这门新的几何学,他自己谨慎地称之为"想象的",因为他还不能指出它的现实意义,虽然他相信会找到这种现实意义的.他的几何的大多数结论对大多数人来说,非但不是"想象的",而且简直是不可想象的和荒诞的.可是无论如何罗巴切夫斯基的思想为几何学的新发展以及各种不同的非欧几里得空间的理论的建立打下了基础.后来这些思想成为广义相对论的基础之一,而且四维非欧几何的一种形式成了广义相对论的数学工具.这样,看来是不可理解的抽象数学体系成了一个最重要的物理理论发展的有力工具.

1.1.4 数学发展简史

大数学家庞加莱说:"若想预见数学的将来,正确的方法是研究它的历史和现状".法国人类学家斯特劳斯说:"如果他不知道他来自何处,那就没有人知道他去向何方."我们需要知道,我们现在何处,我们是如何到达这里的,我们将去何方.数学史将告诉我们来自何处.

数学的发展史大致可以分为四个基本上本质不同的阶段.

第一个时期——数学形成时期.这是人类建立最基本的数学概念的时期.人类从数数开始逐渐建立了自然数的概念,简单的计算法,并认识了最简单的几何形式,逐步地形成了理论与证明之间的逻辑关系的"纯粹"数学.算术与几何还没有分开,彼此紧密地交错着.

第二个时期称为初等数学,即常数数学的时期.这个时期的基本的、最简单的成果构成现在中学数学的主要内容.这个时期从公元前5世纪开始,也许更早一些,直到17世纪,大约持续了两千年.在这个时期逐渐形成了初等数学的主要分支:算术、几何、代数、三角.

按照历史条件不同,可以把初等数学史分为三个不同时期:希腊的、东方的和欧洲文艺复兴时代的.

希腊时期正好与希腊文化普遍繁荣的时代一致.到公元前3世纪,在最伟大的古代几何学家欧几里得、阿基米德、阿波罗尼奥斯的时代达到了顶峰,而终止于公元6世纪.当时最光辉的著作是欧几里得的《几何原本》.尽管这部书是两千多年以前写成的,但是它的一般内容和叙述的特征,却与我们现在通用的几何教科书非常相近.

希腊人不仅发展了初等几何,并把它导向完整的体系,还得到许多非常重要的结果.例如,他们研究了圆锥曲线:椭圆、双曲线、抛物线;证明了某些属于射影几何的定理,以天文学的需要为指南建立了球面几何,以及三角学的原理,并计算出最初的正弦表,确定了许多复杂图形的面积和体积.

在算术与代数方面,希腊人也做了不少工作.他们奠定了数论的基础,并研究丢番图方程,他们发现了无理数,找到了求平方根、立方根的方法,知道算术级数与几何级数的性质.

在几何方面希腊人已接近"高等数学".阿基米德在计算面积与体积时已接近积分运算,阿波罗尼奥斯关于圆锥曲线的研究接近于解析几何.

应该指出,当时我国的算术和代数已达到很高的水平.在公元前

2世纪到1世纪已有了三元一次联立方程组的解法.同时在历史上第一次利用负数,并且叙述了对负数进行运算的规则,也找到了求平方根与立方根的方法.

随着希腊科学的终结,在欧洲出现了科学萧条,数学发展的中心移到了印度、中亚细亚和阿拉伯国家.在这些地方从5世纪到15世纪的一千年中间,数学主要由于计算的需要,特别是由于天文学的需要而得到发展.印度人发明了现代记数法,引进了负数,并把正数与负数的对立和财产与债务的对立及直线上两个方向的对立联系了起来.他们开始像运用有理数一样运用无理数,他们给出了表示各种代数运算包括求根运算的符号.由于他们没有对无理数与有理数的区别感到困惑,从而为代数打开了真正的发展道路.

"代数"这个词本身起源于9世纪的数学家和天文学家穆罕默德·花拉子米.花拉子米的著作基本上建立了解方程的方法.从这时起,求方程的解作为代数的基本特征被长期保持了下来.他的代数著作在数学史上起了重大作用,因为这部作品后来被翻译成拉丁语,曾长期作为欧洲主要的教科书.

中亚细亚的数学家们找到了求根和一系列方程的近似解的方法,找到了"牛顿二项式定理"的普遍公式,他们有力地推进了三角学,把它建成一个系统,并造出非常准确的正弦表.这时中国科学的成就开始传入邻国.约在公元6世纪我国已经会解简单的不定方程,知道几何中的近似计算以及三次方程的近似解法.

到16世纪,所缺少的主要是对数及虚数,还缺乏字母符号系统.正像在远古时代,为了运用整数,应该制定表示它们的符号一样,现在为了运用任意数并对它们给出一般运算规则,就应该制定类似的符号.这个任务从希腊时代就开始而直到17世纪才完成,在笛卡儿和其他人的工作中最后形成了现代符号系统.

在科学复兴时期,欧洲人向阿拉伯学习,并且根据阿拉伯文的翻译熟识了希腊科学.从阿拉伯沿袭过来的印度记数法逐渐在欧洲确

定了下来.

只是到了 16 世纪,欧洲科学终于越过了先人的成就.例如意大利人塔尔塔利亚和费拉里在一般形式上先解了三次方程,然后解了四次方程.在这个时期第一次开始运用虚数.现代的代数符号也制造出来了,其中不仅出现了表示未知数的字母符号,也出现了表示已知数的字母符号;这是韦达在 1591 年作出的.

最后,英国的纳皮尔发明了供天文学作参考的对数,并在 1614 年发表.布利格算出第一批十进对数表是在 1624 年.

当时在欧洲也出现了"组合论"和"牛顿二项式定理"的普遍公式;级数知道得更早,所以初等代数的建立是完成了,以后则是向高等数学,即变量数学的过渡.但是初等数学仍在发展,仍有很多新结果出现.

第三个时期是变量数学的时期.

到 16 世纪,封建制度开始消亡,资本主义开始发展并兴盛起来.在这一时期中,家庭手工业,手工业作坊逐渐地改革为工场手工业生产,并进而转化为以使用机器为主的大工业.因此,对数学提出了新的要求.这时,对运动的研究变成了自然科学的中心问题.实践的需要和各门科学本身的发展使自然科学转向对运动的研究,对各种变化过程和各种变化着的量之间的依赖关系的研究.

作为变化着的量的一般性质和它们之间依赖关系的反映,在数学中产生了变量和函数的概念.数学对象的这种根本扩展决定了数学向新的阶段,即向变量数学时期的过渡.

数学中专门研究函数的领域叫做数学分析,或者叫无穷小分析.这后一名词的来源是,因为无穷小量概念是研究函数的重要工具.

所以,从 17 世纪开始的数学的新时期——变量数学时期可以定义为数学分析出现与发展的时期.

变量数学建立的第一个决定性步骤出现在 1637 年笛卡儿的著作《几何学》.这本书奠定了解析几何的基础,它一出现,变量就进入

了数学,从而运动进入了数学.恩格斯指出:"数学中的转折点是笛卡儿的变数.有了变数,运动进入了数学,有了变数,辩证法进入了数学"(恩格斯《自然辩证法》,人民出版社 1971 年版 236 页).在这个转折以前,数学中占统治地位的是常量,而这之后,数学转向研究变量了.

在《几何学》里,笛卡儿给出了字母符号的代数和解析几何原理,这就是引进坐标系和利用坐标方法把具有两个未知数的任意代数方程看成平面上的一条曲线.解析几何给出了回答如下问题的可能:

1)通过计算来解决作图问题;

2)求由某种几何性质给定的曲线的方程;

3)利用代数方法证明新的几何定理;

4)反过来,从几何方面来看代数方程.

因此,解析几何是这样一个数学部门,即在采用坐标法的同时,用代数方法研究几何对象.

在笛卡儿之前,从古代起在数学中起优势作用的是几何学.笛卡儿把数学引向另一途径,这就是使代数获得更重大的意义.

变量数学发展的第二个决定性步骤是牛顿和莱布尼茨在 17 世纪后半叶建立了微积分.事实上牛顿和莱布尼茨只是把许多数学家都参加过的巨大准备工作完成了,它的原理却要溯源于古代希腊人所创造的求面积和体积的方法.

微积分的起源主要来自两方面的问题:一是力学的一些新问题,已知路程对时间的关系求速度,及已知速度对时间的关系求路程;一是几何学的一些相当古老的问题,作曲线的切线和确定面积和体积等问题.这些问题在古代就研究过,在 17 世纪初期开普勒、卡瓦列里和许多其他数学家也研究过,但是这两类问题之间的显著关系的发现,解决这些问题的一般方法的形成,要归功于牛顿和莱布尼茨.微积分的发现在科学史上具有决定性的意义.

除了变量与函数概念以外，以后形成的极限概念也是微积分以及进一步发展的整个分析的基础.

同微积分一道，还产生了分析的另外一部分：级数理论、微分方程论、微分几何.所有这些理论都是因为力学、物理学和技术问题的需要而产生并向前发展的.

微分方程论是研究这样一种方程，方程中的未知项不是数，而是函数.微分几何是关于曲线和曲面的一般理论.在19世纪还产生了另一个重要分支，即复变函数论，它使分析的内容更加充实.复变函数是将实分析的方法推广到复数域中去了.

分析蓬勃地发展着，它不仅成为数学的中心和主要部分，而且还渗入到数学较古老的范围，如代数、几何与数论.

通过分析及其变量、函数和极限等概念.运动、变化等思想，使辩证法渗入了全部数学.同样地，基本上通过分析，数学才在自然科学和技术的发展中成为精确地表述它们的规律和解决它们问题的得力工具.

在希腊人那里，数学基本上就是几何；在牛顿以后，数学基本上就是分析了.

当然，分析不能包括数学全部；在几何、代数和数论中都保留着它们特有的问题和方法.比如，在17世纪，与解析几何同时还产生了射影几何，而纯粹几何方法在射影几何中占统治地位.

这时还产生了另一个重要的数学部门——概率论.它研究大量"随机"现象的规律问题，给出了研究出现于偶然性中的必然性的数学方法.

在希腊几何的历史上，欧几里得所做的严格和系统的叙述结束了以前发展的漫长道路.和这种情况相似，随着分析的发展必然引起更好地论证理论，使理论系统化，批判地审查理论的基础等这样一些任务，这些任务是19世纪中叶到来的.这项重要而困难的工作由于许多杰出学者的努力而胜利完成了，特别是获得了实数、变量、函数、

极限、连续等根本概念的严格定义.

理论原则的建立是其发展的总结,但不是它的终结,相反地,正是新理论的起点.分析的情形也是这样,由于它的基础的准确化产生了新的数学理论,这就是 19 世纪 70 年代德国数学家康托尔所建立的集合论.在此基础上又产生了分析的一个新部门 —— 实变函数论.同时集合论的一般思想渗入到数学的所有部门.这种"集合论观点"与数学发展的新阶段不可分割地联系在一起.

第四个时期为现代数学.

数学发展的第一时期与第二时期所获得的主要成果,即初等数学中的主要成果已经成为中小学教育的内容.

第三个时期的基本结果,如解析几何(已部分地放入中学)微积分、微分方程,高等代数、概率论等已成为高等学校理工科教育的主要内容.这个时期的数学的基本思想和结论已广泛地为大众所知,几乎所有的工程师和自然科学工作者都或多或少地运用着这些结果.近几十年来,数学应用的状况发生着深刻的变化.这些成果逐渐渗透到社会科学研究的各个领域.因而这些内容的一部分已进入文科各系的教学内容.

与此相反,数学发展的最近阶段,即现代阶段的思想和结果基本上还只是为在数学、力学、物理学及一些新技术领域中工作的科学工作者所使用.

现在在转向叙述数学发展最新阶段的一般特征时,我们只试图简略地给出数学的这些新分支的最一般的特征.

数学发展的现代阶段的开端,以其所有基础部门 —— 代数、几何、分析 —— 中的深刻变化为特征.

还在 19 世纪上半叶,罗巴切夫斯基和波尔约就已经建立了新的非欧几何学,它的思想是别开生面的和出乎意外的.正是从这个时候起,开始了几何学的原则上的新发展,改变了几何学是什么的本来理解.它的研究对象与使用范围迅速扩大.1854 年著名的德国数学家

黎曼继罗巴切夫斯基之后在这个方向上完成了最重要的步骤.他提出了几何学家能够研究的"空间"的种类有无限多的一般思想,并指出这种空间的可能的现实意义.如果说,以前几何学只研究物质世界的空间形式,那么现在,现实世界的某些其它形式,由于它们与空间形式类似,也成了几何学的研究对象,可采用几何学的各种方法对它们进行研究.因此,"空间"这一术语在数学中获得了新的更广泛的,也是更专门的意义,同时几何学方法本身也大大地丰富和多样化了.欧几里得几何本身也发生了很大的变化.现在可研究复杂得多的图形,乃至任意点集的性质.同样地出现了研究图形本身的崭新的方法,在这些研究的基础上,产生了各种新而又新的"空间"和它们的"几何":罗巴切夫斯基空间,射影空间,各种不同维数的欧氏空间、黎曼空间、拓扑空间等,所有这些概念都找到了自己的应用.

在19世纪,代数也出现了质的变化.以往的代数是关于数字的算术运算的学说.这种算术运算是脱离了给定的具体数字在一般形态上形式地加以考察的.也就是说,在代数中,凡量都以字母来表示,按照一定的法则对这些字母进行运算.现代代数在保持这种基础的同时,又把它大大地推广了.现代代数中还考察比数具有更普遍得多的性质的"量",并且研究对这些量的运算,这些运算在某种程度上按其形式的性质来说与加、减、乘、除等普通算术运算是类似的.向量是最简单的例子,我们知道,向量按照平行四边形法则相加.在现代代数中进行的推广达到这样的程度,以致"量"这个术语本身也常常失去意义,而一般地是讨论"对象"了,对这种"对象"可以进行与普通代数运算相似的运算.例如,两个相继进行的运动相当于某一个总的运动,一个公式的两种代数变换相当于一个总的变换等等.与此相应就可讨论运动与变换所特有的"加法".现代代数在一般抽象形式上研究所有这种类似的运算.

现代代数理论是19世纪前半叶从许多数学家的研究中形成的,其中尤以法国数学家伽罗瓦的工作著名.现代代数的概念、方法和结

果在分析,几何、物理以及结晶学中都有重大应用.群论与线性代数是现代代数中内容丰富的两个分支,并在自己的发展中得到很广的应用.

分析也发生了深刻的变化.首先,它的基础得到了精确化,特别是得到了它的基本概念:函数、极限、积分,最后是变量概念本身的精确和普遍定义,实数的严格定义也给出了.这些工作是由一批杰出的数学家完成的,其中有捷克数学家波尔查诺,法国数学家柯西,德国数学家魏尔斯特拉斯、戴德金等.我们还必须提到德国数学家康托尔的集合论.它促进了数学的其他许多新分支的发展,对数学发展的一般进程产生了深刻的影响.

在分析中发展出一系列新的分支,如实变函数论,函数逼近论,微分方程定性理论,积分方程论,泛函分析.在分析和数学物理发展的基础上同几何与代数新思想相结合产生的泛函分析在现代数学中起着特殊重要的作用.

集合论还导致了数学领域的另一分支——数理逻辑的发展.一方面,数理逻辑溯源于数学的起源和基础,另一方面它又和计算技术的最新课题紧密相连.数理逻辑得到了许多深刻的结果.这些结果从一般认识论的观点看来也十分重要.

1.1.5 现代数学发展的新趋向

数学的现代发展不仅表现在现代数学的新领域和高层次中,而且还表现在数学向一切学科与社会部门的渗透和应用.现代数学正在向复杂性进军,研究的对象越来越复杂,其主要表现有以下几方面.

1) 从单变量到多变量,从低维到高维.

2) 从线性到非线性.

3) 从局部到整体,从简单到复杂.

4) 从连续到间断,从稳定到分岔.

5) 从精确到模糊.

6）计算机的使用.

1.1.6　计算机的影响

电子计算机的出现是20世纪科学的最大成就之一，它冲击，影响和促进着现代数学的发展，改变着数学学科本身的特点和面貌.

电子计算机强大的计算能力使数学如虎添翼，比以前任何时候都更有威力和渗透力.一些复杂的数学问题，过去由于求解困难或计算量过大而不易处理和应用，现在可以依靠计算机直接给出数值答案，这不但极大地扩展了数学的应用范围，也改变了对数学求解的概念.

计算机也改变了数学应用的实践方式.天文学中的超新星爆发过程，地学中的地壳运动等都难以进行实验，但却可以用计算机通过数学模型来模拟，从而对各种理论进行检验.这样，科学研究除了传统的理论工作和实验外，又出现了计算机上进行的数学实验.因为这种方法既快又省，所以它具有极大的优越性.

计算机给数学理论提出了一系列新课题.如符号计算，机器证明，人工智能等，这些新课题的研究将扩大计算机的功能，从而进一步解放人的大脑.计算机为数学研究提供了新方法，例如，四色问题这一著名的难题正是借助计算机来解决的.计算机结束了长期以来数学家的工具只是纸和笔，而进入了数学成果的机器生产新时代.

1.1.7　关于中等教育

为了在21世纪为我国培养一大批杰出的科学家，中学数学教育起着关键的作用.以下几点应当受到注意：

1）将应试教育转为素养教育.要培养学生善于思考，有独创精神，而不只是长于记忆，巧于应考.这对我们民族的长远利益是极关重要的.

2）中学数学教育的中心应实现三个转变：从具体数学到概念化数学的转变，发展符号意识；从常量数学到变量数学的转变；从直观

描述到严格证明的转变.建立严密的逻辑思维意识.

3）向学生提供数学主流的核心部分、为学习微积分,统计学和计算机作好准备.

4）计算机教育应尽早进行.建立在计算机与人脑思维相结合之上的新教学法,将有利于培养学生的洞察力、理解力,以及数学直观.

科学和技术已经达到影响人类生活的所有方面的地步,数学也就成为教育议事日程上极其重要的问题.数学是科学和技术的基础.数学在决定国家的各级人才的实力方面起着日益重要的作用.

现在是21世纪初叶,新的世纪已经到来.人类社会已经经历了原始社会、农业社会、工业社会,正在向信息化社会迈进.现在是人类科学技术空前繁荣的时代.我们已经乘上高速列车向未来飞速前进.信息时代,高技术时代就是数学时代.联合国教科文组织在里约热内卢宣言中宣布:“2000—— 世界数学年”.这充分体现了国际社会对数学在社会发展中所具有的重要意义的共识.

在这新旧世纪交替之际,我们应当回顾过去,展望未来,准备好自己,迎接新的挑战.

第二章　数　系

整数是全部数学的基础.

H.闵可夫斯基

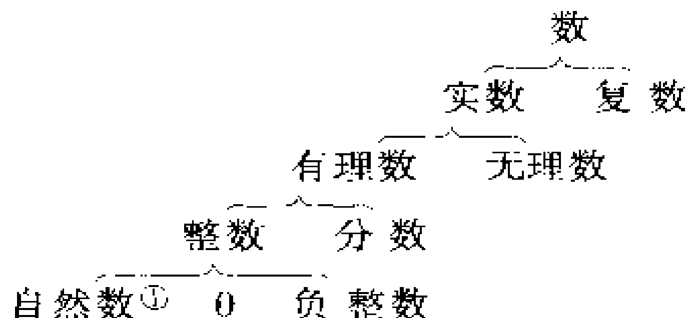
如果不知道远溯古希腊各代所建立和发展的概念、方法和结果,我们就不可能理解近五十年来数学的目标,也不可能理解它的成就.

H.魏尔

数学是研究空间形式和数量关系的科学.数与形是数学中最基本、最原始的概念.从数的概念出发发展出了算术,代数以及庞大的分析系统.从形的概念出发,发展出欧氏几何,非欧几何,直到现代的流形理论.数和形这两个概念在数学的发展过程中,互相影响,交互作用.本讲座就围绕这两个基本概念展开.

数系部分包含三节:1) 无理数的诞生,2) 无限的比较,3) 复数.

通过三节的讲授,我们将建立数系:



注　1) 我们将自下而上地建立数系,指出数的更深刻的对称性,使数的知识精确化.

2) 对无限进行比较是数学发展史上的一次飞跃,也是人类认识史上的一次飞跃,对现代数学的发展具有深刻的影响,我们将从中学知识开始讲到许多深刻的数学思想.

---

① 这里的自然数是指正整数,不包括零.

3）复数与实数一起构成现代数学的基础，是本讲座的必要知识.

§2.1 无理数的诞生

2.1.1 自然数

微积分的基础是实数论，实数的基础是有理数，有理数的基础是自然数. 要真正理解现代数学必须回到自然数. 所有的数学命题最终应归结为关于自然数的命题. 这是现代数学基础研究的成果之一. 克罗内克说："上帝创造了自然数，其余的都是人的工作." 这是说，自然数为稳固的数学结构提供了基础，数学的一切研究从此开始.

研究自然数遇到的第一个问题是记数法与进位制的问题. 目前我们采用的是十进位制，方便清楚地将一切自然数表示了出来. 十进位制是中国人的一大发明. 在商代中期的甲骨文中已有十进位，其中最大的数为三万. 在公元前 4 世纪的《墨经》中也有记述. 印度最早到 6 世纪末才有十进位制. 但是目前使用的记数法却是印度与阿拉伯人引进的.

记数法与十进制的诞生是自然数发展史上的一次飞跃. 无穷多个自然数可以用有限个符号来驾驭，所有的自然数都可以方便清楚地表示出来.

从自然数一出现，人们对它的研究就没有间断过. 不过，开始人们总是研究特殊数的特殊性质，而且往往带有强烈的神秘色彩.

谈到人类最早对自然数的理性认识，应该从毕达哥拉斯谈起. 毕达哥拉斯（Pythagoras，公元前 572 ～ 公元前 500）是古代著名的哲学家和数学家，是毕达哥拉斯教团的创始人. 这个组织虽然是宗教性的，但是其原则却对柏拉图和亚里士多得的思想发生影响，并促进了数学和理性哲学的发展.

毕达哥拉斯学派对自然数作了多方面的、系统的和深入的考察. 他们不仅把数字看成记数的工具，而且看成神圣、完善、友好、幸运及

邪恶的符号.我们举几个例子.他们认为大于1的奇数象征男性,偶数象征女性.5是第一个男性数与女性数之和,因此象征结婚与结合.他们还发现了完全数.完全数就是等于它的真因数之和的数.如6的真因数是1,2,3,而6 = 1 + 2 + 3.数6就变为完美的象征.下一个完全数是28,28有真因数1,2,4,7,14,而28 = 1 + 2 + 4 + 7 + 14.

他们还发现了亲和数.两个数是亲和的,如果每一个数是另一个数的真因数之和.例如284和220是一对亲和数.因为220的真因数是1,2,4,5,10,11,20,22,44,55,110,其和为284;而284的真因数为1,2,4,71,142,其和为220.后来又增添了神秘的色彩与迷信的成分.当时的人们认为,分别写上这两个数的护符会使两数的佩戴者保持友谊.这种数在魔术、占星学、占卦上都起过重要作用.

他们还研究了形数,其中包括三角形数、正方形数,五边形数等等,这些数被看作是某些几何图形中的点的数目,如图2－1所示.许多关于形数的有趣的定理能以纯几何的方法来证明.例如,

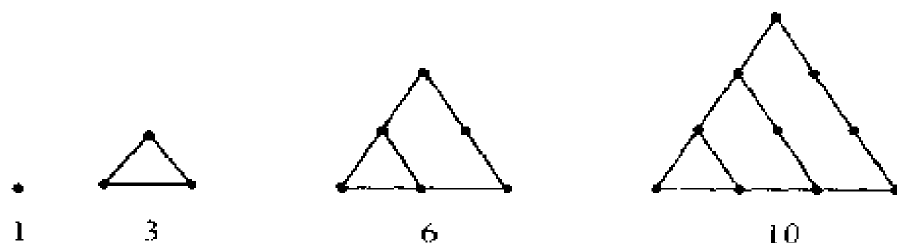


图2－1(a)

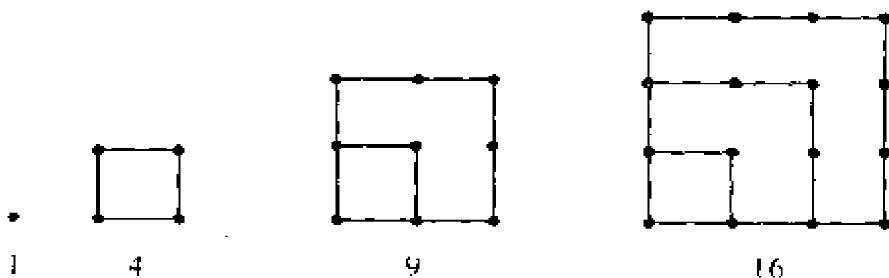


图2－1(b)

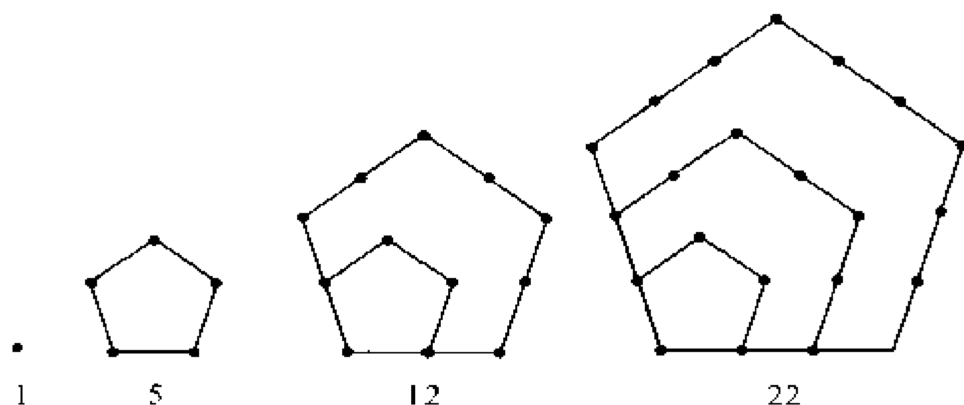


图 2－1(c)

定理 1　任何一个正方形数都是两个相继的三角形数之和.

定理 2　第 n 个五边形数等于第 $n-1$ 个三角形数的三倍加上 n.

定理 3　从 1 开始,任何个相继的奇数之和是完全平方.

证　这三个定理的证明如图 2－2 所示.

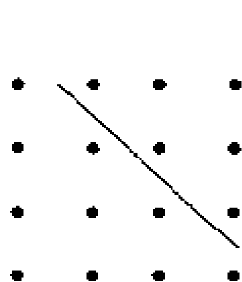


图 2－2(a)

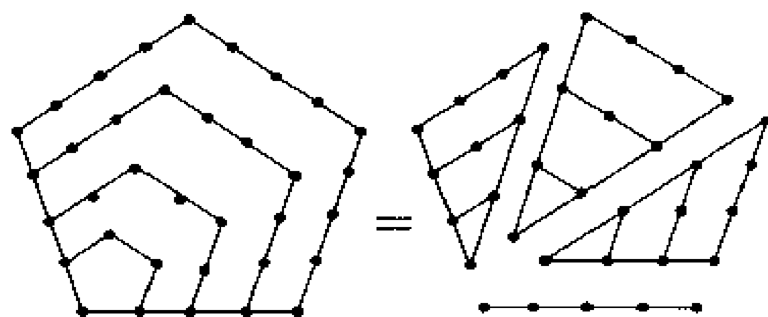


图 2－2(b)

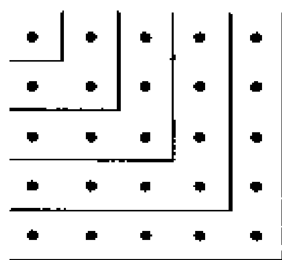


图 2－2(c)

数 7 和 36 在毕达哥拉斯学派里具有特殊的意义.尊敬数 7 是因为巴比伦人给它增添了神秘的意义,又从巴比伦传到了毕达哥拉斯学派.至于数 36,则是它的性质对毕达哥拉斯产生了强烈的印象.36 是自然数数列中前三个数的立方和:$1^3+2^3+3^3$;另一方面,这个数又是自然数列中前四个偶数与前四个奇数之和.按照毕达哥拉斯学派的看法,整个宇宙是建立在前四个偶数与前四个奇数基础之上的.他们认为,用数 36 作的誓言是最可怕的誓言.

毕达哥拉斯学派对周围的生活现象进行了认真而周密的观察,发现了一些奇妙的联系.例如,在悦耳的音乐中毕达哥拉斯学派觉察到了和声的谐音,并注意到在用三根弦发音时,这三根弦的长度之比为 3∶4∶6 时,就得到和声的谐音.他们在其它场合也发现了同样的比例.例如立方体的面数、顶点数、棱数的比,等于 6∶8∶12.在研究同名正多边形覆盖平面的问题时,毕达哥拉斯学派找到了这种覆盖只有三种情况:环绕平面上一个点可以紧密地放 6 个正三角形,或者 4 个正方形,或者 3 个正六边形,见图 2－3.如果注意到这三种情况下正多边形的个数,那么我们可以看到,多边形个数的比为 6∶4∶3.如果我们取这些多边形边数的比,那么它们等于 3∶4∶6.毕达哥拉

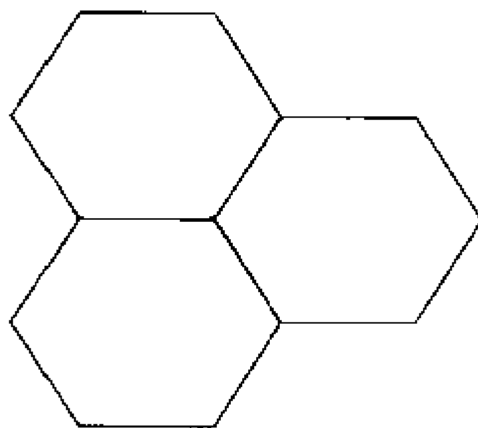


图 2－3(a)

斯学派根据类似的观察更加确信，整个宇宙的现象依附于某种数值的相互关系，也就是存在着"宇宙的和谐".

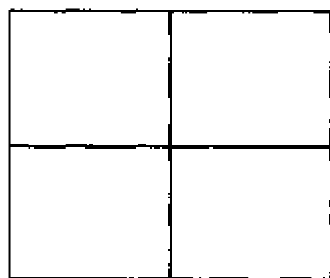


图 2－3(b)

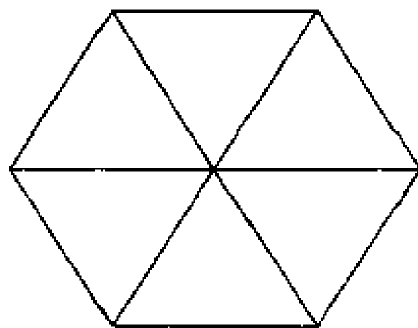


图 2－3(c)

毕达哥拉斯学派研究数学的目的是企图通过揭示数的奥秘来探索宇宙的永恒真理．正像上面指出的，他们发现数与几何图形、数与音乐的和谐，他们还发现数与天体的运行都有密切关系．他们把整个学习过程分成四大部分：1) 数的绝对理论 —— 算术；2) 静止的量 —— 几何；3) 运动的量 —— 天文；4) 数的应用 —— 音乐．合起来称为四艺．后来加上文法、逻辑和修辞，合称七艺．

关于数的神秘学说奠定了毕达哥拉斯学派的哲学基础．毕达哥拉斯学派认为，数是现实的基础，是严密性与次序的依据，是在宇宙体系里控制着自然的永恒关系．数，是世界的法则和关系，是主宰生死的力量，是决定一切事物的条件．事物的实质是仿效着数做出来的．

毕达哥拉斯学派的研究具有正反两方面的长远影响：

1) "毕达哥拉斯创立了数学，并把它变成一门高尚的艺术"(公元前 4 世纪，科学史家欧德缪斯语)．

2) 数学占卜的开始．

2.1.2 代数结构的出现

对自然数可实行两种运算：加法和乘法．这是二元运算．对每一

对有序的自然数 a 和 b，存在一个唯一的自然数 c 和 d，叫做 a 与 b 的和及 a 与 b 的积，记为

$$c = a + b, \qquad d = ab,$$

它们遵从算术基本规律：

1) $a + b = b + a$， 加法交换律

2) $ab = ba$， 乘法交换律

3) $a + (b + c) = (a + b) + c$， 加法结合律

4) $(ab)c = a(bc)$， 乘法结合律

5) $a(b + c) = ab + ac$. 加乘分配律

直到 19 世纪早期，代数还被单纯地看作是符号化的算术. 上述五条成为自然数的代数中总成立的性质. 但是，这些性质是符号的. 它们可以用于自然数以外的元素的集合. 因此，这五条性质也可以看作是其它完全不同的元素体系的性质. 它们的推理构成可应用于自然数的代数；显然它们的推理也构成应用于其它体系的代数. 这就是说，许多不同的体系有共同的代数结构. 这五条基本性质可看作是对特殊类型的代数结构的公设. 从这个观点考虑，代数不再束缚于算术上，而成为纯粹形式的演绎研究了.

1830 年左右，上述代数现代观点的萌芽出现在英国数学家皮考克(George Peacock，1791—1858)的著作中. 皮考克是最先研究代数基本原则的人之一. 1830 年，他发表了《代数论著》一书，试图对代数作出堪与欧几里得"原本"比美的逻辑处理. 他赢得"代数的欧几里得"的称号，为抽象代数的诞生开了先河.

2.1.3 逆运算的作用

从根本上讲，是社会生产发展的需要推动了数学的发展. 但是这些推动是通过数学自身矛盾的发展而实现的. 可以说，社会需求是动力，数学的内部矛盾是杠杆，两者缺一不可. 下面我们来看看逆运算在数的扩充中所起的作用.

人类历史上最先发明的数是正整数，或者叫自然数. 它们是

$$1,2,3,4,\cdots$$

两个自然数之和仍是自然数,例如5与7的和是12.但是两个自然数的差,就不一定是自然数了,例如5减7就不再是自然数.为了使减法永远可能,我们需要扩大自然数的集合:每个自然数与负号“-”结合在一起,产生一个负整数,再补充一个新符号“0”,读作“零”.这样我们得到整数的集合:

$$\cdots,-4,-3,-2,-1,0,1,2,3,4,\cdots$$

在整数集合中,加与减的运算总是畅行无阻的:整数集合对加、减运算是封闭的.自然数的集合没有什么有趣的结构,可是一旦加入0和负整数,这个集合就产生了交换群的结构.

两个整数相乘仍是整数,因而在整数集合中乘法也畅行无阻.但是两个整数相除就可能不再是整数,这就引出了有理数的概念.

所有形如m/n的数的集合称为有理数集,其中m、n都是整数,且$n\neq 0$.有理数集中含有全体整数与通常的分数.每个有理数有无穷多个表示法,例如,1可表示为1/1,2/2,3/3,…,再如2/3可表示为4/6,6/9,8/12,….分数m/n称为最简表示,若m与n没有公共素因子.在全体有理数的集合中,加、减、乘、除都可畅行无阻(当然,0不能作除数),因而有理数对四则运算是封闭的.

这里需要指出的是,上面所述的数的扩充过程是从逻辑上讲的.历史的实际过程是,有理数的诞生远在0与负数之前.

但是,有理数的开方不再是有理数,这就引出了无理数的概念.最后,负数的开方引出了复数.数概念的进一步研究放在下面几节.

从上面我们可以看出,逆运算在数的扩充中起着重要的作用.关于逆运算,我们有两条经验:一是,逆运算的运算法则来自正运算,因此比正运算困难;二是,逆运算引出新东西.这两条经验具有普遍意义.

2.1.4　有理数的稠密性

有理数很重要,是人们实际使用的数,是测量长度,面积,体积,

温度，质量等各种量的工具. 当把测量的刻度逐渐加细时，有理点密密麻麻到处都有. 这是一个重要的基本事实，称为有理数的稠密性.

定义 一个数集在数轴上是稠密的是指，在数轴上，每一个不管处于什么位置，也不论是多么小的区间(a,b)中都存在着这个数集中的点.

定理 4 有理点在数轴上是稠密的.

证 根据数集稠密性的定义，要证有理点在数轴上是稠密的，只要证，不管多么小的区间(a,b)，也不管它落在数轴的什么位置上，总有有理数落在(a,b)中.

若$b-a>1$，则区间(a,b)的长度大于两相邻整数间的间距. 因而，不管(a,b)位于数轴的什么地方，总会有一个整数落在(a,b)中. 这样一来，可以假定$b-a<1$. 我们又可假定(a,b)不含任何整数，否则定理已真. 这样(a,b)就落在某个区间$[n,n+1]$中.

不难看出，我们只需证明有理数在$[0,1]$中稠密就行了. 因为加上一个整数n就可以把这一结果推广到任一区间$[n,n+1]$上去了. 有理数在$[0,1]$内的稠密性是容易证明的：

若$b-a>0.1$，则必有$k/10\in(a,b)$，其中$k\in\{1,2,\cdots,9\}$

若$b-a>0.01$，则必有$k/100\in(a,b)$，其中$k\in\{1,2,\cdots,99\}$

……………………

在一般情况下，把n取得足够大，使得$1/n$小于区间(a,b)的长度就可以了；这时总有一个有理数m/n落在该区间上.

2.1.5 有理数域

全体有理数 —— 整数和分数，仍然满足结合律、交换律和分配律. 而且方程

$$a+x=b,\quad ax=b$$

总有解. 换句话说，在有理数的范围内，所有有理运算 —— 加，减，乘，除 —— 可以无限制地进行，而决不会超出这个范围. 这样一个数的集合叫做一个域. 有理数域是我们遇到的第一个数域，后面我们还

会遇到其它的数域. 为以后使用方便,将数域的性质罗列如下:

1) 对于任意两个数,它们的和是唯一确定的.

2) 对于任意两个数,它们的积是唯一确定的.

3) 存在着一个数 0,它具有性质:对于任意 a,均有 $a+0=a$.

4) 对于每一个数 a,均存在一个数 x,适合等式 $a+x=0$.

5) 满足加法交换律 $a+b=b+a$.

6) 满足加法结合律 $(a+b)+c=a+(b+c)$.

7) 满足乘法交换律 $ab=ba$.

8) 满足乘法结合律 $(ab)c=a(bc)$.

9) 乘法对加法适合分配律 $a(b+c)=ab+ac$, $(b+c)a=ba+ca$.

10) 对每一个数 a 以及每一个数 $b\neq 0$,存在唯一的数 x,满足等式 $bx=a$.

数域是抽象代数中的一个基本概念. 本讲座中我们还会碰到新的数域.

问一个问题:从 0 和 1 出发,通过有理运算可否构造出全部有理数?答案是肯定的:可以. 通过加法可以构造出 2,3,4,以至任何自然数 n;通过减法可以得到全体整数;通过除法就可以得出全体有理数了. 以后我们将给出这段话的几何解释.

英国数学家哈代说:"数学家跟画家或诗人一样,也是造型家". 从 0 和 1 出发,通过有理运算构造出全部有理数,是我们见到的第一个数学造型.

有理数域,克服了自然数系的缺陷,相对说来是比较完美的:对四则运算是封闭的,而且具有稠密性. 它为日常测量提供了一个重要的工具. 因此古希腊人曾设想它是同一条无限长直线上的点相对应的、一个从小到大的量的连续排列的长河. 但是这种关于数的连续性的设想,这种算术与几何自然和谐的美妙图象,不久却被希腊人自己证明是完全错误的.

毕达哥拉斯学派发现了勾股定理,不久他的学派就发现了无理数.

2.1.6　第一次数学危机

在古代的数学家看来与有理数对应的点充满了数轴,现在尚未深入了解数轴性质的人也会这样认为.因此,当发现在数轴上存在不与任何有理数对应的一些点时,在当时人们的心理上引起了极大的震惊,这个发现是早期希腊人的重大成就之一.它是在公元前5世纪或6世纪的某一时期由毕达哥拉斯学派的成员首先获得的.这是数学史上的一个里程碑.毕达哥拉斯学派发现,没有任何有理数与数轴

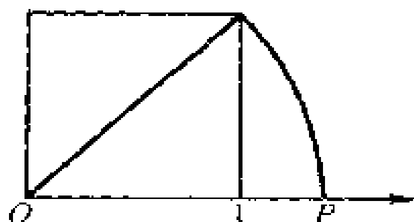


图2-4

上的这样一点相对应(图2-4):距离OP的长度,它等于边长为1的正方形的对角线长.后来,又发现数轴上还存在许多点也不对应于任何有理数.因此,必须发明一些新的数,使之与这样的点相对应;因为这些数不能是有理数,所以把它们称为无理数.

根据勾股定理,边长为1的正方形的对角线其长度为$\sqrt{2}$.为了证明点P不能由一个有理数表示,只须证明$\sqrt{2}$是无理数即可.

定理5　$\sqrt{2}$是无理数.

证明要用算术基本定理(其证明见第四讲).

算术基本定理　设$a>1$是任一正整数,则

$$a=p_1p_2\cdots p_s,$$

其中$p_j(1\leqslant j\leqslant s)$是素数,在不计次序的意义下,表示式是唯一的.

定理5的证明　今用反证法证明,即假定定理的结论不成立,从而引出一个矛盾.现在假设$\sqrt{2}$不是无理数,而是有理数,则$\sqrt{2}=p/q$,其中p和q是整数,且没有公共素因数.于是

$$p = q\sqrt{2},$$

平方得

$$p^2 = 2q^2.$$

它断定，p 的平方是 q 的平方的两倍.我们必需证明一个自然数的平方决不可能是另一个自然数的平方的两倍(不包括 $0^2 = 2 \times 0^2$).

设 p 表示为 d 个素因数的乘积：

$$p = p_1 p_2 \cdots p_d$$

q 表示为 e 个素因数的乘积：

$$q = q_1 q_2 \cdots q_e$$

于是，p^2 是 $2d$ 个素数的乘积，q^2 是 $2e$ 个素数的乘积.这样，$2q^2$ 是 $2e + 1$ 个素数的乘积，因为它有额外的素因数 2.

p^2 分解成素数的数目是偶数，$2q^2$ 分解成素数的数目是奇数，这违反了算术基本定理.这个矛盾证明了不存在平方是 2 的有理数.$\sqrt{2}$ 是有理数导致了矛盾.因此必须放弃这个假设.定理证毕.

注意，在推理中使用的数字 2 的唯一的性质是 2 是素数，因此，同样的推理可以证明：

定理 6　任何素数的平方根都是无理数.

新的情况是 $\sqrt{6}$ 如何?6 既不是素数也不是一个自然数的平方，但 $\sqrt{6}$ 仍是无理数.要证明 $\sqrt{6}$ 是无理数，简单模仿上面的推理是行不通的，因为

p^2	$=$	6	q^2
偶数个素数的乘积		两个素数的乘积	偶数个素数的乘积
(无矛盾)			

因而，需对推理稍作修改.我们不去计算素数的总数，而仅仅计算 p^2 和 $6q^2$ 分解成素数时 2 的数目.

p^2　：　素因数分解中，有偶数个 2(或 0 个 2)

q^2　：　素因数分解中，有偶数个 2(或 0 个 2)

6　：　素因数分解中，有 1 个 2

可见，在 $6q^2$ 的素因数分解中，有奇数个 2. 这个矛盾证明 $\sqrt{6}$ 是无理数.

沿着这个改进的推理前进，可以确定哪些自然数有有理数的平方根，哪些自然数有无理数的平方根.

定理 7 如果在自然数 A 的素因数分解中，至少有一个素数出现奇数次，那么 $\sqrt{A}$ 是无理数.

例如，$\sqrt{6}$，$\sqrt{8}$，$\sqrt{22}$，…，等都是无理数.

定理 8 如果在自然数 A 的素因数分解中，每一个素数都出现偶数次，那么 $\sqrt{A}$ 是有理数.

例如，$\sqrt{4}$，$\sqrt{16}$，… 等都是有理数.

无理数的发现推翻了早期希腊人坚持的另一信念：给定任何两个线段，必定能找到第三个线段，也许很短，使得给定的线段都是这个线段的整数倍. 事实上，即使现代人也会这样认为，如果他还不知道情况并非如此的话. 现在我们取一个正方形，设它的边长为 s，对角线长为 d，并知道 $d = s\sqrt{2}$. 取定这两个线段；如果存在第三个线段 t，使得 s 和 d 都包含 t 的整数倍，我们就有 $s = qt$，$d = pt$，这里 p，q 是整数. 由 $d = s\sqrt{2}$ 得 $pt = qt\sqrt{2}$，从而 $p = q\sqrt{2}$，即 $\sqrt{2} = p/q$，这是一个有理数，与定理 1 相矛盾. 这说明存在不可公度的线段，即不具有共同度量的线段.

2.1.7 历史意义

第一次数学危机表明，当时希腊的数学已经发展到这样的阶段：

1）数学已由经验科学变为演绎科学.

2）把证明引入了数学.

3）演绎的思考首先出现在几何学中，而不是在代数学中，使几何具有更加重要的地位. 这种状态一直保持到笛卡儿解析几何的诞生.

中国、埃及，巴比伦，印度等国的数学没有经历这样的危机，因而一直停留在实验科学，即算术的阶段. 希腊则走上了完全不同的道路，形成了欧几里得的《几何原本》与亚里士多得的逻辑体系，而成

为现代科学的始祖.

在当时的所有民族中为什么只有希腊人认为几何事实必需通过合乎逻辑的论证而不能通过实验来建立?这个原因被称为希腊的奥秘,这个奥秘值得探索.

总之,第一次数学危机是人类文明史上的重大事件.

2.1.8 第一次危机的消除

无理数与不可公度量的发现在毕达哥拉斯学派内部引起了极大的震动.首先,这是对毕达哥拉斯哲学思想的核心,即"万物皆依赖于整数"的致命一击;既然像$\sqrt{2}$这样的无理数不能写成两个整数之比,那么它究竟怎样依赖于整数呢?其次,这与通常的直觉相矛盾,因为人们在直观上总是认为任何两个线段都是可公度的.而毕达哥拉斯学派的比例和相似形的全部理论都是建立在这一假设之上的,突然之间基础坍塌了,已经确立的几何学的大部分内容必须抛弃,因为它们的证明失效了.数学基础的严重危机爆发了.这个"逻辑上的丑闻"是如此可怕,以致毕达哥拉斯学派对此严守秘密.据说米太旁登的希帕苏斯把这个秘密泄露了出去,结果被抛进大海.还有一种说法是,将他逐出学派,并为他立了一个墓碑,说他已经死了.

这个"逻辑上的丑闻"是数学基础的第一次危机,既不容易,也不能很快地被消除.大约在公元前370年,才华横溢的希腊数学家欧多克索斯以及柏拉图和毕达哥拉斯的学生阿契塔给出两个比相等的定义,从而巧妙地消除了这一"丑闻".他们给出的定义与所涉及的量是否可公度无关.其实这也是自然的,因为两个线段的比本来与第三个线段无关.当然从理论上彻底克服这一危机还有待于现代实数理论的建立.在实数理论中,无理数可以定义为有理数的极限,这样又恢复了毕达哥拉斯的"万物皆依赖于整数"的思想.

2.1.9 层次

数学概念是以某种方式按层次安排的,每一层次的概念自身相

互之间以及与属于上、下层次的概念之间被一些复杂的关系联系着.自上而下,层次越深,思想就越深刻,一般说来也更困难.无理数概念比有理数概念深刻.下面将研究更深层次的概念.

2.1.10　反证法

在证明$\sqrt{2}$是无理数的时候我们用了反证法.反证法是数学中经常使用的一种方法,需要给出进一步的说明,以帮助读者逐渐掌握这一证明方法.我们知道,一个定理,或者一个命题,可能是正确的,也可能是错误的.因此,要想知道一个命题正确与否就需要加以证明.但是,有些数学命题给出直接证明是很困难的,而用反证法证明要简捷或容易得多.有些定理,至今除了反证法以外还不能给出其它证明.甚至有这样的定理,它可以用反证法证明,但由于这个定理本身的特点,即使在原则上也不可能给出直接的构造性证明.我们再先举一个使用反证法的简单实例.

例如,有科技书、外文书、文艺书共10本,证明:在这三种书籍中,至少有一种书籍至少有四本.

假如我们采用直接证法,就要把这三种共十本书籍中每种书出现的各种可能都毫无遗漏地考虑到.我们把各种情况列成下表:

科技书	10	9	9	8	8	8	7	7	7	7	…	0	…	0
外文书	0	0	1	0	1	2	0	1	2	3	…	0	…	10
文艺书	0	1	0	2	1	0	3	2	1	0	…	10	…	0

这样就需要列出一个具有66种可能情况的表.从这个表中可以看出,不管出现哪种情况,这十本书中至少有一种书不少于四本,这样就证明了这个命题.

但是,这种直接证明的方法是多么费事啊!幸亏是十本书,三大类.如果是一百本书九大类,甚至更多的书,更多的类,岂不是更复杂吗?

所以这种命题就不适宜用直接证法.我们用反证法来证明.假定命题的结论不对,即每种书籍至多有三本,那么这些书籍的总和将最

多是九本.这与已给条件矛盾,命题就这样被证明了.

应用反证法证明一个命题时,一般往往采用如下的步骤:

1) 假定命题的结论不成立.

2) 进行一系列的推理.

3) 在推理过程中出现了下列情况中的一种:

(1) 与已知条件矛盾;

(2) 与公理矛盾;

(3) 与已知定理矛盾.

4) 由于上述矛盾的出现,可以断言,原来的假定"结论不成立"是错误的.

5) 肯定原来命题的结论是正确的.

总之,用反证法证明命题实际上是这样一个思维过程:我们假定"结论不成立",结论一不成立就会出现毛病,这个毛病是通过与已知条件矛盾,或者与公理、定理矛盾的方式暴露出来的.这个毛病是怎么造成的呢?我们的推理没有错误,已知条件,已知公理、定理没有错误,这样,唯一有错误的地方就是一开始假定的"结论不成立"有错误."结论不成立"与"结论成立"必然有一个正确,既然"结论不成立"有错误,就肯定结论必然成立了.

反证法也称为归谬法.著名的英国数学家 G.H. 哈代(Hardy,1877—1947)对于这种证明方法作过一个很好的评论.在棋类比赛中,经常采用一种策略是"弃子取势"—— 牺牲一些棋子以换取优势.哈代指出,归谬法是远比任何棋术更为高超的一种策略:棋手可以牺牲的是几个棋子,而数学家可以牺牲的却是整个一盘棋.归谬法就是作为一种可以想像的最了不起的一种策略而产生的.

习 题

1. 求证 $\sqrt{33}$ 是无理数.

2. 若 p,q 是奇数,则方程 $x^2+px+q=0$.

(a) 不可能有等根;(b) 不可能有整根.

3. 证明素数的个数是无限的.

§2.2　无限的比较

不管学过数学的人,还是没有学过数学的人,脑子里都有一个无限大的概念.如像无限的时间,无限的空间,直线上有无穷多点,0除1是无穷大,等等.所以,确实存在着一些无穷大的数,再如,"所有整数的个数","一条直线上所有的点的个数",显然它们都是无穷大的.关于这些数,除了说它们是无穷大外,我们能比较两个无穷大的数吗?"所有整数的个数"和"一条直线上所有点的个数"究竟哪个大些?

乍看之下,上面的问题使人发蒙:这问题有意义吗?两个无穷大的数能比较吗?但是,德国著名数学家G·康托尔(Georg Cantor 1845—1918)认真思考了这个问题.

一一对应的概念一直是计数有限集合的根据.康托尔从1874年起发表了一系列重要文章,应用这一概念来计数无限集合,从而产生了关于超限数的重要理论.这在数学发展史上是一个重要的里程碑.

2.2.1　一段富有启发性的历史对话

伽利略曾用意大利文写了两部有名的专论.一部是关于天文学的,题目是"关于托勒密和哥白尼两大世界体系的对话"(The Two Chief Systems,1632),论述了托勒密和哥白尼的宇宙观.另一部是关于物理学的,题目是《关于两种新科学的对话》(The Two New Sciences,1638).在两部著作中都采用了三个文艺复兴时期的绅士对话的形式:见广识多的科学家萨尔维阿蒂,才智犀利的普通人沙格列陀和正统的亚里士多德派辛普利邱.下面转载第二部书中的一段对话,从这段对话可以看出伽利略已认识到无限大的某些性质.让我们

仔细地把谈话的论点检查一下，确定哪些观点是正确的，哪些观点是错误的，然后提出康托尔的基本发现.

辛普利邱："现在有一个我解决不了的难题.很清楚，由于我们可以有一条比另一条线段更长的线段，其中每一条都包含着无穷数目的点，所以我们就不得不承认，对一条线段和线段内的所有点来说，我们有比无限还要大的东西，因为长线段上的无限的点比短线段上的无限的点要多.这种赋予一个无限的数量以大于无限的值的做法使我无法理解."

萨尔维阿蒂："这是当我们企图以有限的智力讨论无限，并赋予它我们给有限的东西同样的性质时所出现的困难.但是我认为这样做是错误的，因为我们对一个无限的量不能说它大于，小于或等于另一个无限的量.要证明这一点，我进行了推理，为了清楚起见，我将以向提出这种困难的辛普利邱提问的形式叙述这个问题.我认为你当然知道哪些数是平方数，而哪些数不是."

辛普利邱："我当然知道一个平方数是由某一个数自乘后得到的；从而 4,9 等数是平方数，它们是由 2、3 等数自乘后得到的."

萨尔维阿蒂："很好，而你也知道乘积叫做平方数，而因子叫做根；另一方面，由两个不相等的因子组成的数不是平方数.因此，我说包括平方数和非平方数在内的所有数比单独的平方数多，对不对？"

辛普利邱："当然是这样".

萨尔维阿蒂："如果我进一步问究竟有多少平方数，一个可能的答案是，存在着和对应的根一样多的平方数，因为每个平方数都有它自己的根，同时每个平方数只有一个根".

辛普利邱："正是这样".

萨尔维阿蒂："但是，如果我追问究竟有多少个根，那么不能否认，有多少个数就有多少个根，因为每个数都是某个平方数的根.这就使得我们必须这样假定：有多少个数就有多少个平方数 …… 不得不承认，平方数和所有的数的总和一样多."

沙格列陀："在这种情况下，人们必须得出什么结论呢？"

萨尔维阿蒂："我认为到现在为止，我们只能这样猜想：所有的数的总数是无限的，平方数的数目是无限的，··· 平方数的数目既不少于所有数的总数，而后者也不多于前者．最后，对无限来说，'等于'，'大于'，'小于' 等属性是不能使用的，它能使用于有限的数量．"

····

上面是一段很精采的对话，它说明远在康托尔的集合论创始之前，伽利略已经对无限有了很好的理解．这段对话较长，这里作一简要总结：

1）任何线段都包含无限个点．

2）线段[0,2]包含线段[0,1]，因而线段[0,2]比线段[0,1]含更多的点．出现了比无限还大的量，辛普利邱不能理解．

3）萨尔维阿蒂证明了，自然数和它的平方数可建立一一对应：

$$\begin{array}{cccccc} 1 & 2 & 3 & \cdots & n & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \cdots & \updownarrow & \cdots \\ 1 & 4 & 9 & \cdots & n^2 & \cdots \end{array}$$

4）萨尔维阿蒂认为，"等于"，"大于"，"小于"等属性不能使用于无限的数量．因而他在一开始就指出，不能以有限的智力来讨论无限．

还有一部分谈话，这里没有摘录．在那段谈话里萨尔维阿蒂承认自己有一个问题还解决不了．这个问题是

5）他还找不出[0,1]区间的点与全体整数的一一对应．

2.2.2 对谈话的分析和解答

现在我们来分析这段对话，弄清楚其中哪些是正确的，哪些是错误的，并由此引出康托尔的理论．我们需要集合论的最初步的知识，这些都是读者在初等数学中已经熟知的东西．

结论1）很容易回答．设$[a,b]$是任意一个区间，其中$a<b$都是

实数.易见,点集

$$a+(b-a)\cdot\frac{k}{n}(n=1,2,3,\cdots.0\leqslant k\leqslant n).$$

的数目是无限的,它们都在区间$[a,b]$中.

设 A 与 B 为两个有限集,自然会发生下面的问题:它们所含元素的个数是否相同.我们可以数一下每一集所含元素的个数是多少,从所得的数字是否相同就可以解决这个问题.但是不数也可以解决问题.例如

$$A=\{a,b,c,d,e\},$$
$$B=\{\alpha,\beta,\gamma,\delta,\varepsilon\}.$$

如果我们细察下面的表:

A	a	b	c	d	e
B	α	β	γ	δ	ε

我们虽然不数,也晓得 A 与 B 的元素个数是相同的.

上面所用的比较法有这样一个特征:对于一个集的每一个元素,另一个集中有一个并且只有一个元素和它对应,反之亦然.这个比较法的优点是它也可以用于无限集.例如,设 N 为自然数全体的集,而 M 为所有形如 $1/n$ 的数全体,用对应法,将 N 中的 n 对应于 M 中的 $1/n$:

N	1	2	3	4	…
M	1	1/2	1/3	1/4	…

立即可以看到 N 与 M 所含元素是一对一对地配得起来的.

现在我们给配对无余的概念以精确的定义:

定义 设 A 与 B 为二集.具有下面性质 φ:使 A 的任一元素 a,有 B 的唯一元素 b 与之对应,并且使 B 的任一元素 b,也有 A 的唯一元素 a 与之对应,此时称 φ 建立了 A 与 B 的一对一的对应(简称一一对应).

定义 若 A 与 B 间能建立一对一的对应,则称 A 与 B 是“对等”

的，或者称它们的势是相同的. 此事记作

$$A \sim B.$$

不难明白，两个有限集只有当它们的元素的个数是相同时才是对等的. 由上可见，"其势相同"一语乃是有限集元素"个数相同"的直接扩充.

下面举几个对等集的例子.

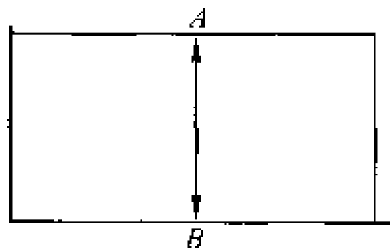


图 2 - 5

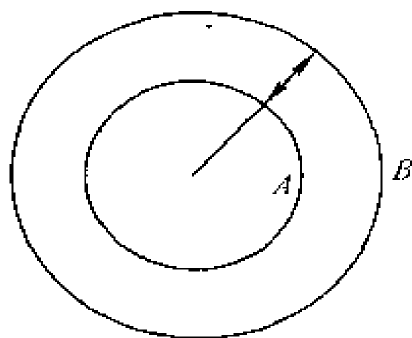


图 2 - 6

设 A 与 B 是一个长方形的一对平行边上点的集(图 2 - 5)，则 $A \sim B$.

设 A 与 B 是两个同心圆周上点的集(图 2 - 6)，显然 $A \sim B$.

所要注意的，此时若将此二圆周展开为直线，则此二线段的长并不相同. 这个例子告诉我们，一个较长的线段并不含有"更多的"点，这种现象，由下例更为显然. 假设 A 表示直角三角形斜边上点的集，B 表示底边上点的集，那么由图 2 - 7，可以看到 $A \sim B$，虽然底边的长小于斜边. 如果我们将底边覆盖在斜边的上面，那么 B 就成为 A 的子集，并且是 A 的真子集(B 是 A 的真子集乃是：$B \subset A$，但 $B \neq A$).

此外我们再举一个例子.

设 N 表示自然数全体的集，而 M 为偶数的全体：

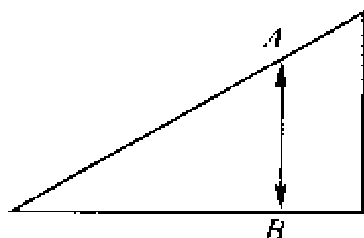


图 2 7

$$N = \{n\}, M = \{2n\}.$$

将此二集用下法使成一一对应：

N	1	2	3	4	…
M	2	4	6	8	…

则 M 与 N 是对等的，虽然 M 是 N 的真子集．因此得到："自然数有多少，偶数也有多少"．

从而辛普利邱的论断"由长线段上的点组成的无限要大于短线段上的点组成的无限"就不正确了，因为这两个集合是对等的，虽然一个集合是另一个集合的真子集．

这个现象似乎是奇怪的，因为它不可能在日常生活中发生．

不难检查，集合$\{1,2,3,4\}$不与它的任何一个真子集对等．这说明有限集与无限集间存在着本质上的差别．

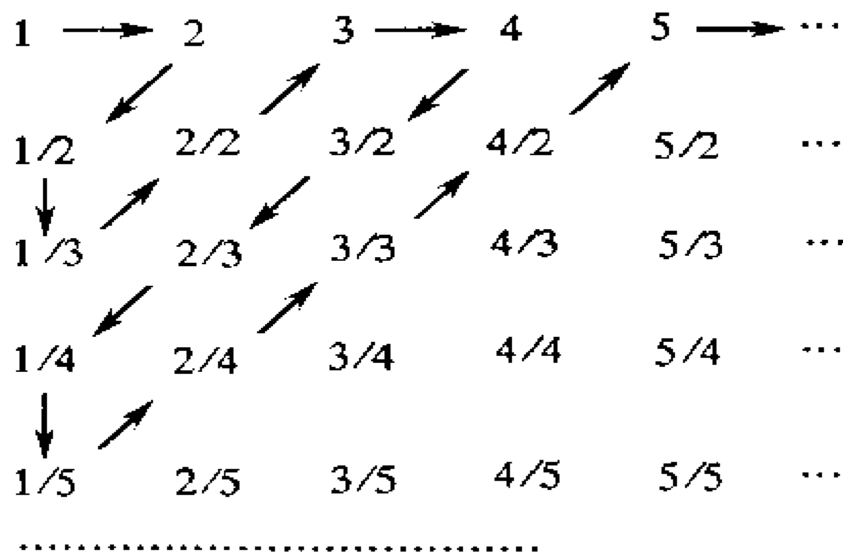
3.2.3 有理数集是可数的

定义 凡与集 N 对等的集 A 都叫作可数集，或称集 A 是可数的．

定理 1 正有理数的集合是可数的．

证 因为每个有理数可以写成 a/b 的形式，这里 a,b 都是整数，所以我们可以把全体正有理数用下面的方阵排出来，使得 a/b 在第 a 列第 b 行．

所有的正有理数都可以按照这样的方法排起来：从 1 开始沿水平线向右走到下一个位置 2，作为序列的第 2 个数．然后沿斜线向下



走到第 1 列的 1/2,作为序列的第 3 个数.再垂直向下走到 1/3,作为序列的第 4 个数.⋯ 如图所示的那样走下去,我们得到序列

1,2,1/2,1/3,2/2,3,4,3/2,2/3,1/4,1/5,⋯.

在这个序列中消掉所有的分数 a/b,其中 a,b 有公因子.于是每个正有理数 r 作为最简分数只在上面的序列中出现一次.这样一来,我们得到序列

1,2,1/2,1/3,3,4,3/2,2/3,1/4,1/5,⋯.

这个序列中包含每一个正有理数,并且只包含一次.这就证明了,正有理数的集合是可数的.

定理 2 一个有限集和一个可数集如无公共元素,那么它们的和集是可数集.

证 设 A 为有限集,B 为可数集:

$$A = \{a_1, a_2, \cdots, a_n\}$$
$$B = \{b_1, b_2, \cdots\}.$$

它们的和集 S 可以表示为

$$S = \{a_1, a_2, \cdots, a_n, b_1, b_2, \cdots\}.$$

在重新编号后可看出,S 是一个可数集.

不难看出，在假定中去掉无公共元素一语，定理仍然正确. 下面几个定理中，这个条件也可去掉.

定理 3 两两不相交的有限个可数集的和集是可数的.

证 我们只对三个被加集的情况加以证明，由此可看出论断的一般性. 设 A, B, C 是三个可数集：

$$A = \{a_1, a_2, a_3, \cdots\},$$

$$B = \{b_1, b_2 b_3, \cdots\},$$

$$C = \{c_1, c_2 c_3, \cdots\},$$

那么它们的和集 S 可以写成

$$S = \{a_1, b_1 c_1, a_2, b_2, c_2, a_3, \cdots\}.$$

所以 S 是可数的.

系 1 全体整数的集合可数的.

因为全体整数的集合是正整数的集合，负整数的集合和只含 0 的集合的三个集合的和集.

系 2 全体有理数的集合可数的.

因为全体有理数的集合是正有理数的集合，负有理数的集合和只含 0 的集合的三个集合的和集.

定理 4 两两不相交的可数个有限集的和集是可数的.

证 设 $A_k (k = 1, 2, 3, \cdots)$ 是两两不相交的可数个有限集：

$$A_1 = \{a_1, a_2, \cdots, a_k\}$$

$$A_2 = \{b_1, b_2, \cdots, b_m\}$$

$$A_3 = \{c_1, c_2, \cdots, c_n\}$$

$$\cdots\cdots\cdots\cdots$$

设 S 是它们的和集. 要证 S 是可数的，只要能把 S 中的元素进行编号就行了. 可将 S 中的元素给以如下的排列：先写出 A_1 中的所有元素，然后写出 A_2 中的元素，如此继续，就可将 S 中的元素依次排出来. 由此可知，S 是可数的.

定理 5 两两不相交的可数个可数集的和集是可数的.

证 把这些集合排列在平面上,每个集合占一行.然后按定理1的路线给元素编号.

上面两个定理及其系似乎使我们可以假定,任何一个无限集合都是可数的.但事实远不是这样.康托尔对此作出了意义深远的发现:实数是不可数的.

2.2.4 实数集是不可数的

萨尔维阿蒂说他找不出[0,1]区间的点与全体整数的一一对应.实际上这种对应是不存在的.这是康托尔的一大贡献.下面我们给出康托尔的定理.

定理 6 实数集是不可数的.

证 康托尔是用反证法证明这一定理的.先假定实数集是可数的,把它们排成一序列.然后找出一个实数不在这个序列中,这就出现一个矛盾.这个矛盾指出,实数集是不可数的.

我们只需证 0 与 1 间的实数是不可数的.为了执行上面的证明方案,假定 0 与 1 间的实数是可数的,并把它们排成下表:

第 1 个数 $0.a_{11}a_{12}a_{13}a_{14}\cdots$

第 2 个数 $0.a_{21}a_{22}a_{23}a_{24}\cdots$

第 3 个数 $0.a_{31}a_{32}a_{33}a_{34}\cdots$

……………………………………

第 k 个数 $0.a_{k1}a_{k2}a_{k3}a_{k4}\cdots$

……………………………………

现在我们取一个数 b:

$$b = 0.b_1b_2b_3b_4\cdots, 1 \leqslant b_i \leqslant 9,\ i = 1,2,\cdots.$$

使得 $b_1 \neq a_{11}, b_2 \neq a_{22}, b_3 \neq a_{33}, \cdots, b_k \neq a_{kk}, \cdots$.这样得到的数 b 在 0 与 1 之间,它不同于序列中的第一个数,也不同于序列中的第二个数,…这就是说,它不同于序列中的任何一个数数 b 不可能是

$0.000\cdots=0$,也不可能是 $0.999\cdots=1$. b 一定严格位于 0 与 1 之间.而它不在序列中.这样的数找到了,定理得证.

证明中使用的方法叫做康托尔对角线法,这种证明方法变成了一种模式,在许多证明中都用到它.康托尔第一次证明这个定理是在 1873 年,上面采用的证明是他在 1890 年给出的第二个证明.

现在我们可以指出有理数集和无理数集的一个重要区别了:前者是可数的,后者是不可数的.首先证明无理数集是不可数的.因为,假定无理数集是可数的,那么,根据定理 3,所有有理数集与所有无理数集的和集也应该是可数的.但是这个和集恰恰是全部实数的集合,是一个不可数集.由此看出,无理数过于丰富,它无法与自然数的集合 N 建立一一对应.这就给出了下面定理的一个新证明:

定理 7 存在着无理的实数.

这是我们证明不是所有的数都是有理数的第二种方法.第一种方法依靠算术基本定理.

由此可以断定,无理数在数量上大大超过有理数.尽管有理数在数轴上处处稠密,但与无理数相比不过是沧海一粟.数不胜数的有理数当初是如此丰富,现在在实数集中却突然变得似乎无足轻重了.

2.2.5 代数数

从中学代数中我们已经熟悉了一次方程与二次方程.它们分别有一个根和两个根.三次方程与四次方程分别有三个根和四个根,我们将放在后面讨论.现在考虑一般的 n 次方程:

$$a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n=0. \qquad (1)$$

这里 $a_i(i=1,2,\cdots,n)$ 是实数或复数.如果把根的重数计算在内,不难猜到,n 次方程有 n 个根,这就是著名的代数基本定理.

代数基本定理 n 次方程(1)在复数域中有 n 个根.

证明见代数部分.

现在考虑一切形如(1)的整系数方程,即 $a_0,a_1,\cdots,a_n$ 都是整

数的方程.

定义 一个实数或复数叫做代数数,如果它是某一个整系数方程的根.

例 1) 方程

$$x - n = 0 \quad (n \text{ 是整数})$$

的根是代数数.因此所有整数都是代数数.

2) 方程

$$ax + b = 0 \quad (a, b \text{ 是整数})$$

的根是有理数.因而,所有的有理数是代数数.

3) 方程

$$x^2 - n = 0 \quad (n \text{ 是自然数})$$

的根是$\sqrt{n}$,因而所有自然数的平方根是代数数.

4) 方程

$$x^2 + 1 = 0$$

的根是 i 和 $-$i.因而 i 和 $-$i 是代数数.

仿此,我们可以造出许多代数数.由此看出,代数数的集合远大于有理数的集合.

定义 任何不是代数数的实数叫做超越数.

这里我们对实数又作了一次划分:代数数与超越数.代数数中包括一切整数和有理数.是否每一个实数都是代数数?如果是这样,那就没有超越数了.要证明一个数是超越数,就必须证明这个数不是任何方次的形如(1)的整系数的方程的根.看来比证明$\sqrt{2}$是无理数要困难多了.现在我们用定理4来证明存在超越数.为了做到这一点,只要证明下面的定理就足够了.

定理 8 代数数的集合是可数的.

在证明定理之前先引进一个概念.每一个形如(1)的多项式联系一个正整数:

$$h = |a_0| + |a_1| + \cdots + |a_n| + n,$$

并把它称为多项式的高.

例 方程

$$x^2 + 4x + 4 = 0$$

的高为 $h = 1 + 4 + 4 + 2 = 11$.

方程

$$x^3 + x - 1 = 0$$

的高为 $h = 1 + 1 + 1 + 3 = 6$.

定理 8 的证明 对 h 的任一个给定值,只有有限个以它为高的多项式.而每个这样的方程最多有 n 个根.所以只有有限个代数数是以 h 为高的方程的根.我们把高为 1 的代数数排成一列,然后再排高为 2 的代数数,这样继续下去,就可以把全部代数数排出来.因而根据定理 4,代数数的集合是可数的.

由定理 6 与定理 8 可以得出:

定理 9 存在超越数.

代数数构成一个庞大的集合.所有有理数都是代数数,大量的无理数也属于这一集合.相比之下,超越数就极难得到.欧拉最早猜到超越数存在(即,并非所有实数都是比较驯顺的代数数),1844 年法国数学家刘维尔证明了下述形式的任何一个数都是超越数:

$$\frac{a_1}{10} + \frac{a_2}{10^{2!}} + \frac{a_3}{10^{3!}} + \cdots$$

其中 a_i 是 0 到 9 的任意整数.1874 年,当康托尔开始研究这个问题的时候,林德曼关于 π 是超越数的证明还没有出台,近 10 年后它的证明才问世.这就是说,在康托尔研究无穷论时,人们只发现了非常少的超越数.也许这些超越数只是实数中的例外,而不是常规.

康托尔已习惯于将例外变为常规.在超越数问题上他成功地实现了这一转变.他首先证明了全部代数数的集合是可数的,由此他推出,超越数在数量上大大超过代数数.

这是一个真正引起争论的定理,因为人们毕竟只知道极少数几

个非代数数的存在.而康托尔断言,绝大多数实数是超越数.但他作出这种断言的时候,并没有给出一个具体的超越数!相反地,他只是数区间中的点,并指出,区间中的代数数只占很小一部分.数学史家埃·贝尔说得好:

"点缀在平面上的代数数犹如夜空中的繁星;而沉沉的夜空则由超越数构成."

2.2.6　无限的算术

定理6指出,实数集合与自然数集合不对等.这表明萨尔维阿蒂的观点是不对的,他认为对无限来说,"等于","大于"或"小于"等属性不能使用.现在我们可以根据康托尔的理论给出无限的比较了.

如果集合 A 与集合 B 的某个子集对等,而不与 B 本身对等,那我们就说,集合 A 的数量少于集合 B.

根据定理4,自然数的集合的数量少于实数集合的数量.

康托尔在发现了不是所有的无限集合都是等价的以后,建立了与有限集合的算术相似的推广到无限集合领域的算术.首先,他引进了表示不同大小的无限集合的符号,正像自然数1,2,3,…… 表示不同大小的有限集合一样.

对任何一个可数的集合,康托尔说:"它有 $\aleph_0$ 个元素".符号 $\aleph$(阿列夫)是希伯来字母表中的第一个字母.下标0将 $\aleph_0$ 与康托尔定义的别的'无限大的数' $\aleph_1,\aleph_2,\cdots$ 区别开来.我们将使用字母 $\aleph_0$ 来表示无限大的数中最小的一个.定理1现在可以读作:"存在 $\aleph_0$ 个正有理数";定理5读作:"$\aleph_0$ 乘 $\aleph_0$ 等于 $\aleph_0$";而定理8读作:"存在着 $\aleph_0$ 个代数数".

对任何与实数等价的集合,康托尔说:"它有 c 个元素".

更进一步,康托尔为这些新数定义了加法和乘法.

康托尔关于无限集合的研究完成于19世纪末,为20世纪的数学家提供了一个重要的工具.

2.2.7 结语

这一节简单介绍了无限的理论.其内容超乎常识,较为深奥,研究方法也独具特色.需要作一总结.

1) 一一对应的概念是研究无限集的一个最基本的概念,也是高等数学的一个最基本的概念.无论是代数学,几何学或分析学都离不开这一概念.如近世代数中的同构,拓扑学中的同胚映射,复分析中的保角映射,都以一一对应为其基础.

2) 有理数是可数的;代数数也是可数的;但实数是不可数的;存在超越数.所有这些结果都加深了我们对实数的认识.

3) 构造性证明和存在性证明.证明某类对象的存在性有两种办法:一是构造这类对象的看得见的摸得着的例子,一是去假定这种对象不存在就必然推出矛盾.这后一种方法是间接的,非构造性方法,我们称它为存在性证明.在初等数学中遇到的证明都是构造性证明.如在初等代数中,一个一元二次方程有两个根.我们给出了求根公式;两个根就摆在我们面前.但本节给出的无理数存在的第二个证明(定理7),及超越数存在(定理9)的证明都属于存在性证明.这就是说,我们知道它存在,但不知道它在哪儿也不知道如何把它找出来.

更早给出存在性证明的是希尔伯特.他在研究不变量理论时给出一个存在性的证明.当时曾引起一场轩然大波.克罗内克认为,没有构造就不能算存在.不变量之王果尔丹甚至说:"这不是数学,这是神学."林德曼说:"令人不快,有害,古怪."但希尔伯特坚持这样的观点:只要能证明附于一个概念的属性绝不会引出矛盾,那么就自然确定了这个概念在数学上是存在的,克莱因支持并赞美这种证明,说:"非常简单,在逻辑上是不可抗拒的."

经过这样一场大辩论,人们逐渐地接受了存在性的证明.果尔丹也优雅地退让了:"我自己一直确信,神学也有它的价值."

虽然希尔伯特不是使用存在性证明的第一人,但他却是第一个认识到它们的深刻价值和意义,并以极端漂亮和动人的方式运用它

们的人.正如希尔伯特指出的:"纯粹的存在性证明之价值恰恰在于,通过它们就可以不必去考虑个别的构造,而且各种不同的构造包括于同一个基本思想之下,使得对证明来说是最本质的东西清楚地突显出来;达到思想的简洁和经济,就是存在性证明生存的理由…禁止存在性证明…等于废弃了数学科学."

4) 康托尔的理论是革命性的.

从古希腊时期直到康托尔时代,哲学家和数学家们都只承认"潜无穷"的存在.也就是说,他们能够在如下意义上同意整数集是无穷的:对于整数集中的任何一个数我们都能找到下一个比它更大的整数,但我们决不可能穷举所有整数.例如,可以想象把每一个整数都写在一张纸条上,然后把这些纸条放进一个袋子里,那么,即使地老天荒我们的工作也永远不会终止.

但是,康托尔的前辈们反对"实无穷"的概念——即,他们反对认为这一过程能够结束或袋子能够装满的观点.用高斯的话说:"…我首先反对将无穷量作为一个实体,这在数学中是从来不容许的.所谓无穷,只是一种说话的方式…"

康托尔不同意高斯的观点.他极愿意将这个装有所有整数的袋子看作一个自足的和完整的实体.他不是将"无穷"仅仅看作一种说话的方式而不予考虑.对于康托尔来说,"无穷"是一个应予以高度重视的实实在在的数学概念,值得我们对其进行严格的理性论证.康托尔的工作是划时代的,对现代数学产生了巨大影响.

习　题

1. 证明从可数集中去掉一个有限子集后剩下的集合还是可数的.

2. 证明任何无限集中必含有一个可数子集.

3. 证明可数集的无限子集是可数的.

§2.3 复数

2.3.1 复数的引进

有许多原因使得数的概念必须越出实数域而引进复数. 最早要求应用复数是为了解三次方程. 但是通常为了容易理解起见, 都是从二次方程引入. 实数域没有提供解二次方程的完整理论. 像

$$x^2 = -1$$

这样一个简单方程没有实数解, 因为任何实数的平方不可能是负的.

我们面前有两条路: 或者宣布这个简单的方程没有解, 或者按照我们熟悉的扩充数的概念的途径引进使这个方程有解的数. 历史上曾有不少数学家取第一种态度, 但更多的数学家取开放的态度, 引进了复数. 长期以来, 许多人不接纳复数, 包括一些大数学家. 经过大约一个世纪的怀疑后, 数学家终于接受了它. 在此基础上, 19 世纪又诞生了复变函数论. 复变函数论被誉为 19 世纪最独特的创造, 是抽象数学中最和谐的理论之一.

复数的引进有何重要意义? 我们认为至少有三点:

1) 使代数方程式论成为一个完美的理论. 代数基本定理是整个数学中最重要的定理之一. 它断言, n 次代数方程有 n 个根. 没有复数的诞生, 就没有代数基本定理. 我们也就不知道一个 n 次代数方程会有多少根.

2) 复数成为研究实分析的得力工具. 法国数学家阿达玛说: "实域中两个真理之间的最短路程是通过复域". 例如, 计算积分, 证明代数基本定理, 研究多项式根的分布等都要借助复数.

3) 复数在电学, 流体力学, 弹性力学等领域都有重要应用.

2.3.2 复数的几何表示

大约同时由维塞尔 (Wessel 1745—1818), 阿尔冈 (Argand

1768—1822）和高斯给出了复数的几何表示.使复数的运算从直观角度来看更为自然.这也是使复数在数学和物理中得到广泛应用的重要原因.

数 $z = x + \mathrm{i}y$ 称为复数，其中 x, y 是实数，而 i 是虚单位，$\mathrm{i}^2 = -1$. 实数 x 和 y 分别称为复数 z 的实部和虚部，记为 $x = \mathrm{Re}\, z$，$y = \mathrm{Im}\, z$.

复数的几何解释就是简单地把 $z = x + \mathrm{i}y$ 用平面上具有直角坐标 x, y 的点来表示(图 2－8). z 的实部是它的 x 坐标，虚部是它的 y 坐标.这样的平面称为复平面.这样一来，复数和复平面上的点建立了一一对应. x 轴上的点对应于实数 $z = x + 0\mathrm{i}$，而 y 轴上的点对应于纯虚数 $z = 0 + \mathrm{i}y$.

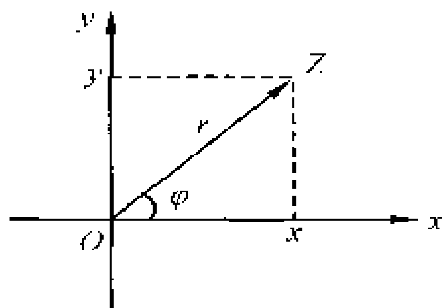


图 2－8

复数可以用向量表示，使复数的实部和虚部分别对应于向量的横坐标与纵坐标.复数的位置也可由极坐标来确定：借助对应于表示复数的向量的长度 r 和这个向量与实轴的夹角 φ 来确定. r 和 φ 分别称为复数 z 的模和辐角，记为

$$r = |z|,\quad \varphi = \mathrm{Arg}\, z.$$

值得注意的是，辐角 Arg z 是多值的，它们之间可以差 2π 的整数倍.通常取由不等式

$$0 \leqslant \mathrm{Arg}\, z < 2\pi$$

所确定的值作为辐角的主值. z 的辐角的主值记为 $\arg z$. 于是我们有

$$\operatorname{Arg} z = \arg z + 2k\pi \quad (k:\text{整数}), \tag{1}$$

这个公式虽然简单却极为重要. 首先,我们要用它去求 n 次单位根,用以研究 n 次单位根作图问题. 其次,与实分析不同,复分析要研究多值函数,复函数的多值性就来源于此.

几个特殊情况:当 z 是正实数时, $\arg z = 0$; z 是负实数时, $\arg z = \pi$; z 是正虚数时, $\arg z = \pi/2$. 特别需要指出的是,0 的辐角是不定的.

2.3.3　复数的三角表示与指数表示

设 $z = x + \mathrm{i}y$. 由辐角和模的定义(图 2 - 8),

$$x = r\cos\varphi, y = r\sin\varphi \tag{2}$$

和

$$r = \sqrt{x^2 + y^2}, \varphi = \arctan\frac{y}{x}. \tag{3}$$

利用(2)可将复数 z 表示为三角形式:

$$z = x + \mathrm{i}y = r(\cos\varphi + \mathrm{i}\sin\varphi).$$

借助欧拉公式

$$\mathrm{e}^{\mathrm{i}\varphi} = \cos\varphi + \mathrm{i}\sin\varphi$$

可将复数 z 化为指数形式

$$z = r\mathrm{e}^{\mathrm{i}\varphi}. \tag{5}$$

这是一种更为紧凑与方便的形式.

2.3.4　复数域

从中学数学中我们已经知道,对复数可以进行加、减、乘、除等四则运算. 复数进行加法运算是实部与实部相加,虚部与虚部相加:

$$(a + b\mathrm{i}) + (c + d\mathrm{i}) = (a + c) + (b + d)\mathrm{i}.$$

其几何意义是,它符合向量加法的平行四边形法则.

复数的减法也类似：

$$(a+b\mathrm{i})-(c+d\mathrm{i})=(a-c)+(b-d)\mathrm{i}.$$

复数作乘法运算没有什么困难，只用到 $\mathrm{i}^2=-1$：

$$(a+b\mathrm{i})(c+d\mathrm{i})=(ac-bd)+(bc+ad)\mathrm{i}.$$

但这种运算很难看出乘法的几何意义. 我们借助复数的指数表示给出其几何意义. 设 $z_1=r_1\mathrm{e}^{\mathrm{i}\theta_1}$，$z_2=r_2\mathrm{e}^{\mathrm{i}\theta_2}$ 则

$$z_1z_2=r_1r_2\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)}.$$

因而，复数乘法是模相乘，辐角相加：

$$|z_1\cdot z_2|=|z_1|\cdot|z_2|,$$

$$\arg(z_1\cdot z_2)=\arg z_1+\arg z_2.$$

除法也类似：

$$\frac{r_1}{r_2}=\frac{r_1\mathrm{e}^{\mathrm{i}\theta_1}}{r_2\mathrm{e}^{\mathrm{i}\theta_2}}=\frac{r_1}{r_2}\mathrm{e}^{\mathrm{i}(\theta_1-\theta_2)}.$$

复数除法是模相除，辐角相减：

$$\left|\frac{z_1}{z_2}\right|=\frac{|z_1|}{|z_2|},$$

$$\arg\frac{z_1}{z_2}=\arg z_1-\arg z_2.$$

复数的运算对加法和乘法仍然满足交换律、结合律和分配律，也就是满足域的基本性质. 因而全体复数构成一个域. 复数域包含实数域为其真子域，因为我们可以把 $a+0\mathrm{i}$ 看成和实数 a 一样.

通过引进符号 i，我们把实数域扩充到了复数域. 在这个域中特殊的二次方程

$$x^2=-1$$

有两个解 $x=\mathrm{i}$ 和 $x=-\mathrm{i}$. 实际我们得到的比这更多. 容易验证，每一个二次方程都有解. 这种方程可写为

$$ax^2+bx+c=0,$$

它的解为 $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. (6)

当 $b^2 - 4ac \geqslant 0$ 时,解为实数. 当 $b^2 - 4ac < 0$ 时,

$$\sqrt{b^2 - 4ac} = \sqrt{-(4ac - b^2)} = \mathrm{i}\sqrt{4ac - b^2},$$

从而解为复数. 有了复数之后,二次方程在任何时候都有两个根(重根算二次). 事实上,我们得到的更多:在复数域中 n 次方程有 n 个根. 这就是代数基本定理所断言的.

2.3.5 乘方与开方

乘方与开方对后面的内容具有重要意义,需要作较为详细的讨论. 我们将得到一些重要公式,希望读者掌握好.

先讲乘方. 设 $z = r(\cos\varphi + \mathrm{i}\sin\varphi)$,由乘法法则,

$$z^2 = r^2(\cos\varphi + \mathrm{i}\sin\varphi)^2 = r^2(\cos 2\varphi + \mathrm{i}\sin 2\varphi),$$

$$z^3 = r^3(\cos\varphi + \mathrm{i}\sin\varphi)^3 = r^3(\cos 3\varphi + \mathrm{i}\sin 3\varphi).$$

一般地,

$$z^n = r^n(\cos n\varphi + \mathrm{i}\sin n\varphi). \tag{7}$$

自然会问,n 是负整数,上面的公式也成立吗?由除法法则,

$$z^{-1} = \frac{1}{z} = \frac{1}{r}(\cos(-\varphi) + \mathrm{i}\sin(-\varphi))$$

$$= \frac{1}{r}(\cos\varphi - \mathrm{i}\sin\varphi),$$

$$z^{-2} = \frac{1}{z^2} = \frac{1}{r^2}(\cos\varphi(-2\varphi) + \mathrm{i}\sin(-2\varphi))$$

$$= \frac{1}{r^2}(\cos 2\varphi - \mathrm{i}\sin 2\varphi).$$

一般地,

$$z^{-n} = r^{-n}(\cos n\varphi - \mathrm{i}\sin n\varphi).$$

所以,不管 n 是正整数还是负整数,(7) 式都成立. 在 $r = 1$ 的特殊情况,我们有棣莫弗(A. De Moivre, 1667 – 1754) 公式:

$$(\cos \varphi + \sin \varphi)^n = \cos n\varphi + \mathrm{i}\sin n\varphi. \tag{8}$$

利用二项式定理,

$$\begin{aligned}(\cos\varphi + \sin\varphi)^n = {} & \cos^n\varphi + n\mathrm{i}\cos^{n-1}\varphi\sin\varphi + \cdots + \\ & n\mathrm{i}^{n-1}\cos\varphi\sin^{n-1}\varphi + \mathrm{i}^n\sin^n\varphi.\end{aligned} \tag{9}$$

把(8),(9)结合起来可以得出通过 $\cos\varphi$, $\sin\varphi$ 表示 $\cos n\varphi$ 和 $\sin n\varphi$ 的公式.例如,当 $n = 2$ 时,

$$\begin{aligned}\cos 2\varphi + \mathrm{i}\sin 2\varphi &= (\cos\varphi + \mathrm{i}\sin\varphi)^2 \\ &= \cos^2\varphi + 2\mathrm{i}\cos\varphi\sin\varphi - \sin^2\varphi,\end{aligned}$$

比较实部和虚部,得

$$\cos 2\varphi = \cos^2\varphi - \sin^2\varphi, \quad \sin 2\varphi = 2\sin\varphi\cos\varphi.$$

当 $n = 3$ 时,

$$\begin{aligned}\cos 3\varphi + \mathrm{i}\sin 3\varphi &= (\cos\varphi + \mathrm{i}\sin\varphi)^3 \\ &= \cos^3\varphi + 3\mathrm{i}\cos^2\varphi\sin\varphi - 3\cos\varphi\sin^2\varphi - \mathrm{i}\sin^3\varphi.\end{aligned}$$

比较实部和虚部,得

$$\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi = 4\cos^3\varphi - 3\cos\varphi. \tag{10}$$

$$\sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi = -4\sin^3\varphi + 3\sin\varphi. \tag{11}$$

注意,复数间的一个等式相当于实数间的两个等式.通过到复数域的旅行,我们得到了实数域的结果. n 越大,这种方法的优越性就越明显.上面的结果是哈达玛的名言"实数域中的两个真理的最短路程是通过复数域"的佐证.

这里特别指出,(10)将在三等分任意角的问题中出现.

下面我们转而研究求方根的问题.设 n 是正整数,a 是任意复数.考虑方程

$$z^n - a = 0 \quad 或 \quad z^n = a, \tag{12}$$

其解为

$$z = \sqrt[n]{a}. \tag{13}$$

为了把 n 个根具体地表达出来,设 $z = \rho(\cos\theta + \mathrm{i}\sin\theta)$, $a = r(\cos\varphi + \mathrm{i}\sin\varphi)$,则

$$\rho^n(\cos n\theta + i\sin n\theta) = r(\cos \varphi + i\sin \varphi).$$

由此得 $\rho^n = r, \quad n\theta = \varphi + 2k\pi \quad (k = \pm 1, \pm 2, \cdots).$

从而 $$\rho = \sqrt[n]{r}, \theta = \frac{\varphi + 2k\pi}{n}.$$

这样一来,

$$\sqrt[n]{a} = \sqrt[n]{r}\left(\cos \frac{\varphi + 2k\pi}{n} + i\sin \frac{\varphi + 2k\pi}{n}\right). \tag{14}$$

令 $k = 0, 1, 2, \cdots, n-1$ 就可得出 n 个不同的根,这里 $\sqrt[n]{r}$ 取的算术值.

公式(14)表明,所有这些不同的根都具有相同的模.相邻两个根的辐角差为 $\frac{2\pi}{n}$.由此可知,任何复数的 n 次方根都有 n 个不同的值,这些值位于半径为 ρ 的圆周上,形成一个正 n 边形的 n 个顶点(图 2 - 9).

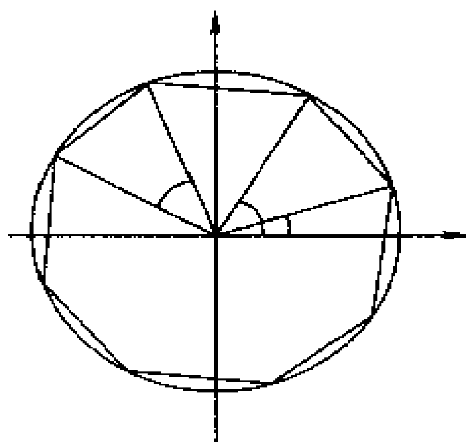


图 2 - 9

2.3.6 单位根

在公式(14)中,令 $a = 1$,就得到 1 的 n 次单位根:

$$\sqrt[n]{1} = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \quad k = 0,1,2,\cdots,n-1.$$

再强调一次，在复数域，1恰有 n 个不同的 n 次方根，它们可以用单位圆的一个内接正 n 边形的顶点来表示(图 2－10). 当 $k = 0$ 时

$$\cos\theta + i\sin\theta = 1$$

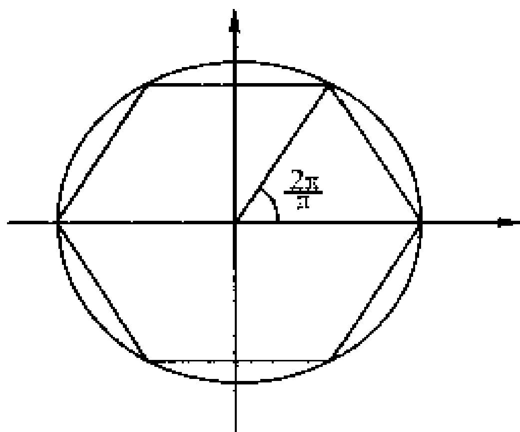


图 2－10

是正多边形的第一个顶点. 正多边形的第二个顶点是

$$\alpha = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}. \tag{15}$$

下一个顶点是 $\alpha \cdot \alpha = \alpha^2$，因为把向量 α 旋转 $2\pi/n$ 就得到它. 再下一个顶点是 α^3、等等. 第 n 步之后，我们回到顶点 1，即我们有

$$\alpha^n = 1.$$

这也可以由棣莫弗公式(8)推出：

$$\left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^n = \cos 2\pi + i\sin 2\pi = 1.$$

可见，α 是方程 $z^n = 1$ 的一个根. 下一个顶点 α^2 也是 $z^n = 1$ 的一个根，因为

$$(\alpha^2)^n = \alpha^{2n} = (\alpha^n)^2 = 1^2 = 1.$$

由此，下述 n 个数

$$1, \alpha, \alpha^2, \cdots, \alpha^{n-1}$$

都是1的 n 次方根.不难看出,这个指数序列再往后,或序列中加上 α 的负指数都不会产生新根.例如

$$\alpha^{-1} = \frac{1}{\alpha} = \frac{\alpha^n}{\alpha} = \alpha^{n-1},$$

$$\alpha^{n+1} = \alpha^n \cdot \alpha = \alpha.$$

它们只是简单地重复以前的值.

如果 n 是偶数,则正 n 边形有一个顶点在 -1 处.

特别地,当 n 是奇素数 p 时,除1外的 $p-1$ 个根均为复数,而且它们都不是次数比 p 低的 n 次单位根,我们称它们为本原单位根.本原单位根在分圆域的理论中很重要.

例　我们给出三次单位根.显然,第一个顶点是 $z=1$.设第二个顶点是 ω,则第三个顶点是 ω^2:

$$\omega = \cos\frac{2}{3}\pi + \mathrm{i}\sin\frac{2}{3}\pi = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2},$$

$$\omega^2 = \cos\frac{4}{3}\pi + \mathrm{i}\sin\frac{4}{3}\pi = -\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}.$$

这两个根后面将用到.

有了单位根就可将任意复数的 n 次方根表示出来.设 a 是任一复数,我们来求 $\sqrt[n]{a}$.只要知道了 $\sqrt[n]{a}$ 的一个根,就可用单位根把其它 $n-1$ 个根求出来.不妨设它的一个根是 β.设 α 是由(15)式给出的1的 n 次单位根,则 $\sqrt[n]{a}$ 的 n 个根是:

$$\beta, \beta\alpha, \beta\alpha^2, \cdots, \beta\alpha^{n-1}.$$

这容易验证.事实上,

$$\beta^n = a, \alpha^n = 1 \Rightarrow (\beta\alpha^k)^n = \beta^n\alpha^{nk} = a \cdot 1 = a.$$

这说明 $\beta\alpha^k (k = 0, 1, 2, \cdots, n-1)$ 是 $\sqrt[n]{a}$ 的根.

例　求 $\sqrt[3]{-8}$ 的三个根.

解　显然,-2 是一个根.其它两个根是

$$-2\omega = 1 - i\sqrt{3}, \ -2\omega^2 = 1 + i\sqrt{3}.$$

下面我们研究方程

$$z^n - 1 = 0.$$

这是一个 n 次方程,但容易化成一个 $n-1$ 次的方程.由

$$z^n - 1 = (z-1)(z^{n-1} + z^{n-2} + \cdots + z + 1),$$

知

$$z^n - 1 = 0,$$

当且仅当 $z = 1$ 或

$$z^{n-1} + z^{n-2} + \cdots + z + 1 = 0 \tag{16}$$

成立.这时(16)必然为 $\alpha, \alpha^2, \cdots, \alpha^{n-1}$ 这些值所满足.

定义 方程(16)叫做分圆方程.

对 n 是素数的情况,高斯详细地考察了分圆方程.他证明了它们总能归结为解一串较低次的方程,并且找到了能用二次根式解出的充要条件.但这个条件的必要性的证明只是到了伽罗瓦才有了严格的基础.

例 解4次分圆方程

$$x^4 + x^3 + x^2 + x + 1 = 0. \tag{17}$$

解 解这个4次方程是为了作正五边形.用 x^2 除(17)的两边,得

$$x^2 + \frac{1}{x^2} + x + \frac{1}{x} + 1 = 0.$$

令 $\omega = x + \dfrac{1}{x}$,由于

$$\omega^2 = \left(x + \frac{1}{x}\right)^2 = x^2 + \frac{1}{x^2} + 2,$$

方程(17)可化为

$$\omega^2 + \omega - 1 = 0.$$

这个方程有两个根

$$\omega_1=\frac{-1+\sqrt{5}}{2},\quad \omega_2=\frac{-1-\sqrt{5}}{2}.$$

因而 1 的 5 次复根是下列两个方程的根：

1) $x+\frac{1}{x}=\omega_1$ 或 $x^2+\frac{1}{2}(1-\sqrt{5})x+1=0$,

2) $x+\frac{1}{x}=\omega_2$ 或 $x^2+\frac{1}{2}(1+\sqrt{5})x+1=0$.

利用二次方程的求根公式，可以把这两个方程的四个复根都求出来，并且都可借助开方运算将它们表示出来．后面我们将证明，这些数都是可构造数、可用直尺圆规构造出来．因而正五边形可用直尺圆规作图．

最后指出，在有理数域中加入 n 次单位根所成的数集仍然构成一个域，叫做分圆域．分圆域在证明费马大定理的过程中起了重要的作用．

2.3.7 复数的确认

16 世纪的数学家对负数还持怀疑态度，负数的平方根当然更是荒谬绝伦．意大利数学家卡尔达诺(Gerolamo Cardano 1501—1576)对三次方程的解法作出重大贡献．他在解三次方程的过程中几次用到复数，但最终他还是把它们放弃了，因为“它们既模不透，又没有用途”．大约经过了一代人的时间，意大利数学家邦贝利(Bombelli Rafael 1526—1572)迈出了勇敢的一步．他把虚数看成是运载数学家从实系数三次方程到达其实数解的必要工具．这就是说，从熟悉的实数域出发，最终回到实数解，但中途不得不进入我们所不熟悉的虚数世界，以完成我们的旅行．所以复数是从三次方程而不是从二次方程获得原动力的．并由此得到无可争辩的合法地位．

但在 16，17 世纪，人们对复数仍有极大的疑虑．这清楚地反映在莱布尼茨的一段话：

“神灵在分析的奇观中找到了超凡的显示，这就是那个理想世界

的征兆.那个介于存在与不存在的两栖物,我们称之为虚数的(−1)的平方根".

这个疑虑到了高斯才得以消除.

习　题

1. 已知正三角形的两个顶点为 a,b,试写出一切可能下的另一顶点.

2. 已知正方形的两个顶点为 a,b 试写出一切可能下的另外两顶点.

3. 求出五次单位根的具体表达式.

4. 设 z_1,z_2,z_3 是适合下列条件的三点:

$$z_1+z_2+z_3=0;|z_1|=|z_2|=|z_3|=1.$$

求证 z_1,z_2,z_3 是内接于单位圆的正三角形的三个顶点.

第三章　连分数及其在天文学上的应用

数学既不严峻，也不遥远，它既和几乎所有的人类活动有关，又对每个真心感兴趣的人有益.

R. C. Buck

三代以上，人人皆知天文.‘七月流火’，农夫之词也；‘三星在天’，妇人之语也；‘月离于毕’，戍卒之作也；‘龙尾伏辰’，儿童之谣也.后世文人学士，有问之而茫然不知者矣.

顾炎武《日知录》卷三十

本章讲述连分数的初步概念，并给出它的一个重要而有趣的应用——在天文学上的应用.最后讲连分数的性质，并引出有趣的斐波那契级数.

§3.1　从祖冲之的圆周率谈起

3.1.1　辗转相除法

读者从小学数学中就已经熟悉了求两个正整数的最大公约数的辗转相除法.由于这一方法对本章内容的基本性与重要性，我们需要在这里作一回顾.

以下将两个正整数 a,b 的最大公约数记为 (a,b).给定两个正整数 a 和 b，并设 $a \geqslant b$.用 b 除 a 得商 a_0，余数 r，写成式子，

$$a = a_0 b + r, 0 \leqslant r < b, \tag{1}$$

这是最基本的式子.若 $r = 0$，则 b 可除尽 a，a 与 b 的最大公约数就是 b.

若 $r \neq 0$，再用 r 除 b，得商 a_1，余数 r_1，即

$$b = a_1 r + r_1, \quad 0 \leqslant r_1 < r. \tag{2}$$

如果 $r_1 = 0$，那么 r 除尽 b，由(1)也除尽 a．又，任何一个除尽 a 和 b 的数，由(1)也一定除尽 r．因此 $r = (a, b)$．

如果 $r_1 \neq 0$，则用 r_1 除 r，得商 a_2，余数 r_2，即

$$r = a_2 r_1 + r_2, \quad 0 \leqslant r_2 \leqslant r_1. \tag{3}$$

如果 $r_2 = 0$，那么由(2)，r_1 是 b 和 r 的公约数，由(1)它也是 a 和 b 的公约数．反之，若一个数能整除 a 和 b，那么由(1)它一定能除尽 b 和 r，由(2)它一定除得尽 r, r_1，所以 r_1 是 a, b 的最大公约数．

如果 $r_2 \neq 0$，再用 r_2 除 r_1，如法进行．因此用辗转相除法得出一系列方程：

$$a = a_0 b + r, \quad 0 \leqslant r < b,$$
$$b = a_1 r + r_1, \quad 0 \leqslant r_1 < r,$$
$$r = a_2 r_1 + r_2, \quad 0 \leqslant r_2 < r_1,$$
$$\cdots\cdots$$
$$r_{n-2} = a_n r_{n-1}, \quad 0 = r_n.$$

余数形成一个递减的序列 $b > r_1 > r_2 > \cdots \geqslant 0$．作有限次除法后，就以一个含有余数 $r_n = 0$ 的方程结束了，并且得到

$$(a, b) = r_{n-1}.$$

最大公约数等于最后一个非零除数．

例　求 362 与 450 的最大公约数．

解　$450 = 1 \times 362 + 88$，

$362 = 4 \times 88 + 10$，

$88 = 8 \times 10 + 8$，

$10 = 1 \times 8 + 2$，

$8 = 4 \times 2$．

最后一个非零余数是 2，所以 $(450, 362) = 2$．

3.1.2 祖冲之的约率 22/7 和密率 355/113

中华民族是个伟大的民族.在中华民族的历史上产生过许多杰出的数学家,祖冲之就是其中之一.他生于公元429年,卒于公元500年.他的儿子祖暅和他的孙子祖皓,也都是数学家,善算历.

关于圆周率 π,祖冲之的贡献有二:

(1) $3.1415926 < \pi < 3.1415927$;

(2) 用 22/7 作为约率,355/113 作为密率.

这些结果是刘徽割圆术之后的重要发展.刘徽从圆内接正六边形算起,令边数一倍一倍地增加,即按 12,24,48,96,…,1536,…,的顺序逐次算出六边形、十二边形、…… 的面积,这些数值逐步地逼近圆周率.用这个方法可以无限精密地逼近圆周率,但每一次都比圆周率小.

祖冲之的结果(1) 从上下两个方面给出了圆周率的误差范围.这个事实容易看出,不必多讲.下面我们将详细讲结果(2).从

$$\frac{355}{113} = 3.1415929\cdots$$

看出,355/113 惊人精密地接近圆周率,准确到六位小数.这一发现比欧州人早了一千年.法国人奥托(Valenlinus Otto) 在 1573 年才发现这个分数.有些人认为那时的人们喜欢用分数来计算,这把问题看简单了.其中孕育了不少道理,这道理可用来推算天文上的许多现象.这就无怪乎祖冲之祖孙三代都是算历的专家了.这个约率和密率涉及到"用有理数最佳逼近实数"的问题.

3.1.3 连分数

1) 引言　现在暂将约率、密率的问题放一下,研究一个新问题.利用辗转相除法,可以把一个数,例如$\frac{9}{7}$,可以写成如下的形式:

$$\frac{9}{7}=1+\frac{2}{7}=1+\frac{1}{\frac{7}{2}}=1+\frac{1}{3+\frac{1}{2}}=1+\frac{1}{3+\frac{1}{1+\frac{1}{1}}}.$$

初看起来似乎没有什么比这更简单,更无意义的事情了,其实不然,这种形式的分数对许多数学问题,特别是对研究数的性质问题具有很大启发性,这种分数称为**连分数**.

17 世纪和 18 世纪的许多大数学家都研究过连分数.即使在今天它仍然是一个活跃的课题.

设想一个学代数的学生试图用下面的方法去解二次方程式

$$x^2-3x-1=0. \tag{4}$$

他首先用 x 遍除各项,接着把这方程式写成形式

$$x=3+\frac{1}{x},$$

未知量 x 仍出现在这个方程式的右边,因此可用与它相等的量,即 $3+1/x$ 来代替它.这就给出

$$x=3+\frac{1}{x}=3+\frac{1}{3+\frac{1}{x}}.$$

反复几次用 $3+1/x$ 代替 x,就得到表达式

$$x=3+\frac{1}{3+\frac{1}{3+\frac{1}{3+\frac{1}{3+\frac{1}{3+\frac{1}{x}}}}}}. \tag{5}$$

因为 x 连续在右端这个"多层"分数中出现,它似乎并没有更接近于求出方程式(4)的解.但是让我们更仔细地来考察一下方程式(5)的右端.每进行一步停一次,我们看到,它包含一系列的分数

$$3,3+\frac{1}{3},3+\frac{1}{3+\frac{1}{3}},3+\frac{1}{3+\frac{1}{3+\frac{1}{3}}},\cdots. \tag{6}$$

把它们化为简分数,并进而化为十进小数,便依次得出下面的数

$$3,\frac{10}{3}=3.333\cdots,\frac{33}{10}=3.3,\frac{109}{33}=3.30303\cdots.$$

终于令人惊喜地发现,这些数给出了二次方程式(4)的正根的越来越好的近似值.二次方程式的求根公式指出,这个根实际上等于

$$x=\frac{3+\sqrt{13}}{2}=3.302775\cdots,$$

它约等于3.303,与上面最后一个结果的前三位小数是一致的.

这些初步的计算提出了两个有趣的问题.首先,如果我们算出越来越多的分数(6),是否能不断得到 $x=(3+\sqrt{13})/2$ 的越来越好的近似值呢?其次,假定我们把得出(5)的步骤无限继续下去,以至取代(5)得出一个表达式

$$x=3+\frac{1}{3+\frac{1}{3+\cdots}}, \tag{7}$$

其中三个黑点代表字"等等",并指出接续的分数是无穷无尽的.那么,(7)右边的表达式等于 $(3+\sqrt{13})/2$ 吗?这两个问题的答案都是肯定的.后面将给出解答.

2)简单连分数和它的渐近分数.形如

$$a_1+\frac{b_1}{a_2+\frac{b_2}{a_3+\frac{b_3}{a_4+\cdots}}}$$

的表达式叫做连分数.在一般情况下,$a_1,a_2,a_3,\cdots,b_1,b_2,b_3,\cdots$ 可以是实数或复数,项可以有限,也可以无限.我们将限于讨论简单连

分数.它们取如下的形式

$$a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cdots}},\tag{8}$$

其中第一项 a_1 通常是正的或负的整数,也可以是 0,项 $a_2,a_3,\cdots$ 是正整数.只含有限项的简单连分数叫有限简单连分数,它取形式

$$a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cdots\cfrac{1}{a_{n-1}+\cfrac{1}{a_n}}}},\tag{9}$$

式中仅含有限个项 $a_1,a_2,\cdots,a_n$.一个比(9)简便的记法是

$$a_1+\frac{1}{a_2}+\frac{1}{a_3}+\frac{1}{a_4}+\cdots+\frac{1}{a_n},\tag{10}$$

第一个 + 号之后的 + 号都写低,表示"降了一层".

例 67/29 的连分数是

$$\frac{67}{29}=2+\frac{9}{29}=2+\cfrac{1}{\cfrac{29}{9}}=2+\cfrac{1}{3+\cfrac{1}{\cfrac{9}{2}}}=2+\cfrac{1}{3+\cfrac{1}{4+\cfrac{1}{2}}}$$

或
$$\frac{67}{29}=2+\frac{1}{3}+\frac{1}{4}+\frac{1}{2},\tag{11}$$

展开的过程就是不断地作带余除法,即辗转相除法.我们用通常辗转相除法的格式把计算过程重写如下:

67		29
58	2	27
9	3	2
8	4	2
1	2	0

这样根据计算结果(由中间的数目给出),就又可写出(11).这个例子展示给我们,求一个有理数(或者说分数)的连分数展式的方法就是辗转相除法,并且总在有限步内完成.另一方面,给定一个形如(9)的有限连分数,沿辗转相除法的逆过程算回去,总得到一个有理数.于是我们有

定理 1 任何一个有理数都能展为有限简单连分数;任何一个有限简单连分数都可化为一个有理数.

定理 1 指出,一个无理数的连分数展式将含有无限多项.

例 求$\sqrt{2}$的连分数.

解

$$\sqrt{2}=1+\sqrt{2}-1=1+\cfrac{1}{\cfrac{1}{\sqrt{2}-1}}$$

$$=1+\cfrac{1}{\sqrt{2}+1}=1+\cfrac{1}{2+\sqrt{2}-1}$$

$$=1+\cfrac{1}{2+\cfrac{1}{\cfrac{1}{\sqrt{2}-1}}}=1+\frac{1}{2}+\frac{1}{2}+\cdots$$

用连分数表示实数时,有理数和无理数有了明显的区别:有限连分数表示有理数,而无限连分数表示无理数.这比用十进小数表示实数具有明显的优越性.因为有理数的小数表示可以是无限循环小数.

这里还有一个问题,就是如何把负数展为连分数?

例 把$-\dfrac{37}{44}$展为连分数.

解 由

$$-\frac{37}{44}=-1+1-\frac{37}{44}=-1+\frac{7}{44},$$

只需展$\dfrac{7}{44}$就可以了.

现在继续研究(10)式.我们把数 $a_1, a_2, \cdots, a_n$ 叫做连分式的部分商.利用它们可以构成分数

$$c_1 = a_1, c_2 = a_1 + \frac{1}{a_2}, c_3 = a_1 + \frac{1}{a_2} + \frac{1}{a_3}, \cdots.$$

它们分别由原连分数在第一、第二、第三层,… 处切断而得到,这些分数分别叫做连分数的第一个、第二个、第三个,… 渐近分数.

例 67/29 的前三个渐近分数分别是

$$2,$$

$$2 + \frac{1}{3} = \frac{7}{3} = 2.333\cdots,$$

$$2 + \frac{1}{3} + \frac{1}{4} = \frac{30}{13} = 2.307\cdots.$$

与 67/29 = 2.3103448 比较可看出,它们都是 67/29 的近似值,并且一个比一个更精确.连分数在解决许多有趣的问题中是一个得力的工具.下面我们分别讨论连分数的一些重要应用.

例 求 $\sqrt{11}$ 的近似值.

解 可以用求平方根的方法求 $\sqrt{11}$ 的近似值,但用连分数更方便.我们有

$$3 \leqslant \sqrt{11} \leqslant 4,$$

$$\begin{aligned} \sqrt{11} &= 3 + (\sqrt{11} - 3) \\ &= 3 + \frac{1}{(\sqrt{11} + 3)/2} \\ &= 3 + \frac{1}{3 + (\sqrt{11} - 3)/2} \\ &= 3 + \frac{1}{3} + \frac{1}{\sqrt{11} + 3} \\ &= 3 + \frac{1}{3} + \frac{1}{6 + (\sqrt{11} - 3)}. \end{aligned}$$

重复这一过程就可得到

$$\sqrt{11}=3+\frac{1}{3+}\,\frac{1}{6+}\,\frac{1}{3+}\,\frac{1}{6+}\,\frac{1}{3+}\,\frac{1}{6+(\sqrt{11}-3)}.$$

取渐近分数,不难得出如下的近似值:

$$\sqrt{11}\approx 3,\frac{10}{3},\frac{63}{19},\frac{199}{60},\frac{1257}{379},\frac{3970}{1197},\frac{25077}{7561}.$$

用小数表示,这些近似值依次是

$$\sqrt{11}\approx 3,3.33333333,3.31578947,3.31666666,3.31662269,$$
$$3.31662489,3.31662478.$$

这些值是很精确的近似值.实际上,

$$\sqrt{11}=3.31662479\cdots.$$

例　一个分子分母很大的分数用起来很不方便,连分数可以帮我们找一个分子分母较小的分数来近似它,且误差很小.如10436/43200的连分数展式是

$$\frac{10463}{43200}=\frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{5+}\,\frac{1}{64}$$

具体计算如下:

10463		43200
9436	4	41852
1027	7	1348
963	1	1027
64	3	321
64	5	320
0	64	1

分数的部分渐近分数是

$$\frac{1}{4},\ \frac{1}{4+}\frac{1}{7}=\frac{7}{29},\ \frac{1}{4+}\frac{1}{7+}\frac{1}{1}=\frac{8}{33},$$

$$\frac{1}{4+}\frac{1}{7+}\frac{1}{1+}\frac{1}{3}=\frac{31}{128},\ \frac{1}{4+}\frac{1}{7+}\frac{1}{1+}\frac{1}{3+}\frac{1}{5}=\frac{163}{673}.$$

例 如，8/33 = 0.2424…，而 10463/43200 = 0.242199. 8/33 与 10463/43200 的差别在小数后第 3 位.

3.1.4 约率与密率的内在意义

我们来求圆周率 π 的连分数展式. 取 π 的近似值为3.14159265，并和 1 比较作如下计算：

3.14159265		1
3	3	
0.14159265	7	0.99114855
0.13277175	15	0.00885145
0.00882090	1	0.00882090
		0.00003055

得到

$$\pi=3+\frac{1}{7+}\frac{1}{15+}\frac{1}{1+}\cdots$$

前四个渐近分数为

3 [径一周三，《周髀算经》]，

$3+\frac{1}{7}$ [约率，何承天(公元 429 – 447)]，

$3+\frac{1}{7+}\frac{1}{15}=\frac{333}{106}$，

$3+\frac{1}{7+}\frac{1}{15+}\frac{1}{1}=\frac{355}{113}$ [密率，祖冲之(公元 429 – 500)].

原来约率和密率竟藏在连分数理论的背后！要知道连分数的近代理论开始于 R. 邦贝利(Rafael Bombelli，1526 ~ 1572)，祖冲之去世

一千年后他还没有诞生.那么,祖冲之知道连分数吗?他用什么办法找到的密率?据文献记载,比他略晚的同时代印度数学家阿利亚伯哈塔(ryabhata)的著作中包含了用连分数求线性不定方程的一般解的尝试,他大约死于公元550年.同时代或稍晚,阿拉伯与希腊的著作中也偶有发现.我们可以猜想,祖冲之的思想中已经有了连分数的概念.实际算出来

$$\frac{22}{7} \approx 3.142, \frac{355}{113} \approx 3.1415929.$$

误差分别在小数点后第三位和第七位,这个精确度即使在现在的生产实践中也是够用的.用比 $\pi = 3.14159265$ 更精密的圆周率来计算,我们可以得出

$$\pi = 3 + \frac{1}{7+}\frac{1}{15+}\frac{1}{1+}\frac{1}{292+}\frac{1}{1+}\frac{1}{1+}\cdots.$$

由此我们可以得到 355/113 之后的一个渐近分数是 103993/33102,这是一个很不容易记忆,也不便于应用的数.以下的数据说明,分母比 7 小的分数不比 22/7 更接近于 π,而分母等于 8 的也不比 22/7 更接近于 π.计算见下表.

分母 q	分子 p	$\pi - p/q$
1	3	0.1416
2	6	0.1416
3	9	0.1416
4	13	0.1084
5	16	0.0584
6	19	−0.0251
7	22	−0.0013
8	25	0.0166

我们还可以通过更多的计算指出,在分母不比 114 大的分数中

以 355/113 最接近于 π. 而 22/7 和 355/113 又是两个相当便于记忆和使用的分数.

习　题

1. 把下面的分数化为有限简单连分数: $\frac{17}{11}$, $\frac{51}{33}$, 3.54, 3.14159.
2. 若 $\frac{p}{q} = 3 + \frac{1}{4} + \frac{1}{1} + \frac{1}{5}$, 求 $\frac{p}{q}$.
3. 求 11/17 的有限简单连分数, 并与 17/11 的连分数相比较.
4. 求 $\sqrt{3}$ 的连分数展式.
5. 若 $p > q$, 且 $\frac{p}{q} = \alpha_1 + \frac{1}{\alpha_2} + \frac{1}{\alpha_3} + \cdots + \frac{1}{\alpha_n}$, 则
$$\frac{p}{q} = \frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \cdots + \frac{1}{\alpha_n}.$$

§3.2　连分数在天文学上的应用

3.2.1　为什么四年一闰, 而百年又少一闰

天文学和年代学中的许多问题可以用数论的概念来计算和陈述. 下面我们将讨论几个有趣的天文学问题. 先讨论闰年问题. 如果地球绕太阳一周是 365 天整, 那么我们就不需要分平年与闰年了; 也就是没有必要每隔四年把二月份的 28 天改为 29 天了.

如果地球绕太阳一周恰是 $365\frac{1}{4}$ 天, 那我们每四年加一天的算法就很精确, 没有必要每隔一百年又少加一天了. 如果地球绕太阳一周恰恰是 365.24 天, 那一百年就有 24 个闰年, 即四年一闰而百年少一闰, 这就是我们用的历法的来源. 由 $\frac{1}{4}$ 可知, 每四年加一天; 由 $\frac{24}{100}$

可知,每百年加 24 天.但是,事实并不这样简单.地球绕日一周的时间,即天文年是 365.2422 天.这一小误差逐渐引起了季节和日历关系之间的难以预料的大变动.例如,在 16 世纪,春分是三月十一日,而不是原来的三月廿一日.中国历史上曾经有过多次重大的历法改革,其根本原因就在于此.

现在让我们用求连分数的渐近分数来求得更精密的结果.我们知道地球绕太阳一周需时 365 天 5 小时 48 分 46 秒,也就是

$$365+\frac{5}{24}+\frac{48}{24\times 60}+\frac{46}{24\times 60\times 60}=365\,\frac{10463}{43+60}(\text{天}).$$

将它展为连分数:

$$365\,\frac{10463}{43200}=365+\frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{5+}\,\frac{1}{64}.$$

算法见前.分数的部分渐近分数是

$$\frac{1}{4},\ \frac{1}{4+}\,\frac{1}{7}=\frac{7}{29},\ \frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1}=\frac{8}{33},$$

$$\frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1+}\,\frac{1}{3}=\frac{31}{128},$$

$$\frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{5}=\frac{163}{673},$$

$$\frac{1}{4+}\,\frac{1}{7+}\,\frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{5+}\,\frac{1}{64}=\frac{10436}{43200}.$$

和 π 的渐近分数一样,这些渐近分数也一个比一个精密.这说明,四年加一天是初步的最好的近似值,但 29 年加 7 天更精密些,33 年加 8 天又更精密些,而 99 年加 24 天正是我们百年少一闰的由来.由数据也可晓得,128 年加 31 天更精密.积少成多,如果过了 43200 年,照百年 24 闰的算法,一共加了 $432\times 24=10368$ 天.但是照精密的计算,却应当加 10463 天,这样一来,少加了 95 天.这就是说,按照百年 24 闰的算法,过 43200 年后,人们将提前 95 天过年,也就是秋初就要过

年了!所以历法又规定:世纪数不能被4整除的世纪年如1700,1800,1900,2100,2200,2300,… 等不是闰年,而其余的世纪年如1600,2000,2400,… 等是闰年.但这又显得有点太短了.现在有人提出,4000,8000,… 等不作为闰年.这是一个仍未解决的问题.

3.2.2 公历的改革.

目前世界上大多数国家使用的公历虽然精度比较高,但从实用角度看还存在一些缺点.这些缺点中最明显的有:

1) 一年四季,各季长度不等,有90,91,92天三种.因此上半年与下半年的长度也不相等;

2) 各月的日数不等,有28,29,30,31天四种,大小月安排无规律;

3) 每日的星期数不固定,随年份而变.如1998年的元旦是星期四,1999年的元旦是星期五.

因此,从使用方便,容易记忆这点来讲,公历是不理想的.为了使公历更加完善,1910年在英国伦敦召开了一次国际改历会议,具体讨论公历的改革问题.据统计,到1927年国际上的改历方案就有一百四十多种.很多国家为此设立了专门的改历委员会.从目前来看,比较引人注意的所谓"世界历"方案有两种.一种是一年分12个月的,叫做"十二月世界历".另一种是一年分13个月的"十三个月世界历"(见表).这两种历法都克服了上述缺点,便于记忆,便于使用,且年年相同,永久可用.这样一来,挂历就不必年年印了.缺点是特别列出的闰日和年终国际节都不在日序之内,对记录社会活动和历史事件带来了麻烦.所以到现在还没有诞生一个为大家普遍接受的新公历.

十二月世界历

一季度	1 月					2 月					3 月				
日	1	8	15	22	29		5	12	19	26		3	10	17	24
一	2	9	16	23	30		6	13	20	27		4	11	18	25
二	3	10	17	24	31		7	14	21	28		5	12	19	26
三	4	11	18	25		1	8	15	22	29		6	13	20	27
四	5	12	19	26		2	9	16	23	30		7	14	21	28
五	6	13	20	27		3	10	17	24		1	8	15	22	29
六	7	14	21	28		4	11	18	25		2	9	16	23	30
二季度	4 月					5 月					6 月				
日	1	8	15	22	29		5	12	19	26		3	10	17	24
一	2	9	16	23	30		6	13	20	27		4	11	18	25
二	3	10	17	24	31		7	14	21	28		5	12	19	26
三	4	11	18	25		1	8	15	22	29		6	13	20	27
四	5	12	19	26		2	9	16	23	30		7	14	21	28
五	6	13	20	27		3	10	17	24		1	8	15	22	29
六	7	14	21	28		4	11	18	25		2	9	16	23	30
三季度	7 月					8 月					9 月				
日	1	8	15	22	29		5	12	19	26		3	10	17	24
一	2	9	16	23	30		6	13	20	27		4	11	18	25
二	3	10	17	24	31		7	14	21	28		5	12	19	26
三	4	11	18	25		1	8	15	22	29		6	13	20	27
四	5	12	19	26		2	9	16	23	30		7	14	21	28
五	6	13	20	27		3	10	17	24		1	8	15	22	29
六	7	14	21	28		4	11	18	25		2	9	16	23	30
四季度	10 月					11 月					12 月				
日	1	8	15	22	29		5	12	19	26		3	10	17	24
一	2	9	16	23	30		6	13	20	27		4	11	18	25
二	3	10	17	24	31		7	14	21	28		5	12	19	26
三	4	11	18	25		1	8	15	22	29		6	13	20	27
四	5	12	19	26		2	9	16	23	30		7	14	21	28
五	6	13	20	27		3	10	17	24		1	8	15	22	29
六	7	14	21	28		4	11	18	25		2	9	16	23	30

十三月世界历

月份 星期	1	2	3	4	5	6	7	8	9	10	11	12	13
日	1	1	1	1	1	1	1	1	1	1	1	1	1
一	2	2	2	2	2	2	2	2	2	2	2	2	2
二	3	3	3	3	3	3	3	3	3	3	3	3	3
三	4	4	4	4	4	4	4	4	4	4	4	4	4
四	5	5	5	5	5	5	5	5	5	5	5	5	5
五	6	6	6	6	6	6	6	6	6	6	6	6	6
六	7	7	7	7	7	7	7	7	7	7	7	7	7
日	8	8	8	8	8	8	8	8	8	8	8	8	8
一	9	9	9	9	9	9	9	9	9	9	9	9	9
二	10	10	10	10	10	10	10	10	10	10	10	10	10
三	11	11	11	11	11	11	11	11	11	11	11	11	11
四	12	12	12	12	12	12	12	12	12	12	12	12	12
五	13	13	13	13	13	13	13	13	13	13	13	13	13
六	14	14	14	14	14	14	14	14	14	14	14	14	14
日	15	15	15	15	15	15	15	15	15	15	15	15	15
一	16	16	16	16	16	16	16	16	16	16	16	16	16
二	17	17	17	17	17	17	17	17	17	17	17	17	17
三	18	18	18	18	18	18	18	18	18	18	18	18	18
四	19	19	19	19	19	19	19	19	19	19	19	19	19
五	20	20	20	20	20	20	20	20	20	20	20	20	20
六	21	21	21	21	21	21	21	21	21	21	21	21	21
日	22	22	22	22	22	22	22	22	22	22	22	22	22
一	23	23	23	23	23	23	23	23	23	23	23	23	23
二	24	24	24	24	24	24	24	24	24	24	24	24	24
三	25	25	25	25	25	25	25	25	25	25	25	25	25
四	26	26	26	26	26	26	26	26	26	26	26	26	26
五	27	27	27	27	27	27	27	27	27	27	27	27	27
六	28	28	28	28	28	28①	28	28	28	28	28	28	28②

① 闰年的闰日所在位置

② 年终国际节日

3.2.3 农历的月大月小、闰年闰月

农历的大月三十天、小月二十九天是怎样安排的?

我们先说明什么叫朔望月.出现相同月面所间隔的时间称为朔望月.也就是从满月(望)到下一个满月,从新月(朔)到下一个新月,从蛾眉月(弦)到下一个同样的蛾眉月所间隔的时间.我们把朔望月取作农历月.

已经知道朔望月是 29.5306 天,把小数部分展为连分数

$$0.5306=\frac{1}{1}\,{}_{+}\,\frac{1}{1}\,{}_{+}\,\frac{1}{7}\,{}_{+}\,\frac{1}{1}\,{}_{+}\,\frac{1}{2}\,{}_{+}\,\frac{1}{33}\,{}_{+}\,\frac{1}{1}\,{}_{+}\,\frac{1}{2},$$

它的渐近分数是

$$\frac{1}{1},\frac{1}{2},\frac{8}{15},\frac{9}{17},\frac{26}{49},\frac{867}{1634},\frac{893}{1683}.$$

也就是说,就一个月来说,最近似的是 30 天,两个月就应当一大一小,而 15 个月中应当 8 大 7 小,17 个月中 9 大 8 小等等.就 49 个月来说前两个 17 个月里,都有 9 大 8 小,最后 15 个月里,有 8 大 7 小,这样在 49 个月中,就有 26 个大月.

再谈农历的闰月的算法.地球绕日一周需 365.2422 天,朔望月是 29.5306 天,而它正是我们通用的农历月,因此一年中应该有

$$\frac{365.2422}{29.5306}=12.37\cdots=12\,\frac{10.8750}{29.5306}$$

个农历的月份,也就是多于 12 个月.因此农历有些年是 12 个月;而有些年有 13 个月,称为闰年.把 0.37… 展成连分数

$$0.37\cdots=\frac{1}{2}\,{}_{+}\,\frac{1}{1}\,{}_{+}\,\frac{1}{2}\,{}_{+}\,\frac{1}{2}\,{}_{+}\,\frac{1}{1}\,{}_{+}\,\frac{1}{3},$$

它的渐近分数是

$$\frac{1}{2},\frac{1}{3},\frac{3}{8},\frac{7}{19},\frac{10}{27}.$$

因此,两年一闰太多,三年一闰太少,八年三闰太多,十九年七闰太少.如果算得更精密些

$$\frac{10.8750}{29.5306}=\frac{1}{2}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{1}+\frac{1}{16}+\frac{1}{1}+\frac{1}{5}+\frac{1}{2}+\frac{1}{6}+\frac{1}{2}+\frac{1}{2}.$$

它的渐近分数是

$$\frac{1}{2},\frac{1}{3},\frac{3}{8},\frac{4}{11},\frac{7}{19},\frac{116}{315},\frac{123}{334},\frac{731}{1935},\cdots.$$

这里需要指出，至迟在春秋时代我们的祖先就已经创造了"十九年七闰法"，相当完满地把我们的历法建筑在科学的基础之上，远远走在世界各国的前列.

3.2.4 二十四节气

廿四节气是很多人都熟悉的，尤其在农村，是家喻户晓. 现在我们使用的日历上，在节气那一天都写着"今日立春"、"今日夏至"等字样. 廿四节的名称是：立春、雨水、惊蛰、春分、清明、谷雨、立夏、小满、芒种、夏至、小暑、大暑、立秋、处暑、白露、秋分、寒露、霜降、立冬、小雪、大雪、冬至、小寒、大寒. 为了便于记忆，劳动人民创立了一首歌诀：

春雨惊春清谷天，
夏满芒夏暑相连，
秋处露秋寒霜降，
冬雪雪冬寒又寒.

廿四节气在我国是逐步形成的. 至迟在殷商时代已经有了夏至、冬至等概念，以后逐渐丰富，到了西汉初期已经有了完整的二十四节气. 在我国古代，二十四节气的日期是由测定太阳影子的长度来决定的.《周髀算经》和《后汉书律历志》等许多古书都记载着二十四节气的日影长度数值. 这说明二十四节实际上是太阳视运动的一种反映，与月亮运动毫无关系. 因此二十四节在公历中的日期基本上变化不大，有四句口诀很好记：

公历节气真好算，一月两节不改变．上半年来六、廿一，下半年来八、廿三．

这就是说，节气在上半年的公历日期都在六日和廿一日，而下半年都在八日和廿三日．由于太阳运动的不均匀性，这些日子可能有一、二日的出入，但不会差更多．节气在古代本称为"气"，每个月含有二个气，一般在前的叫"节气"，在后的叫"中气"，后人把节气和中气统称为节气．按照古人的规定，每个月由所含的中气来表征，如含冬至的月就是十一月，含雨水的月就是正月．各月的节气和中气分配如下：

	正月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
节气	立春	惊蛰	清明	立夏	芒种	小暑	立秋	白露	寒露	立冬	大雪	小寒
中气	雨水	春分	谷雨	小满	夏至	大暑	处暑	秋分	霜降	小雪	冬至	大寒

这个表同农历闰月的安排有密切关系．

3.2.5　闰月放在哪儿

农历的闰月究竟怎样安置，历史上曾有过不同的处理．大致上，西汉初期以前，都把闰月放在一年的末尾．例如，汉初把九月作为一年的最后一个月，那时的闰月就放在九月之后，称为"后九月"．到了后来，随着历法的逐步精密，安置闰月的方法也有了新规定，这就是把不含有中气的月份作为闰月，这个置闰规则直到今天仍在使用．为什么农历有的月份会没有中气呢？原来，两个节气或两个中气之间平均日数为 $365.2422 \div 12 = 30.4368$ 日，而一个朔望月是 29.5306 日，两者有将近一天的差数，因此，中气在农历月份中的日期会逐月有将近一天的推迟．这样继续下去，必然有的月份的中气正好落在这个月的最后一天，那么下个月就没有中气了，而是出现在再下一个月的月

初了.按照前节的规定,每个月都有自己固定的中气,那么把没有中气的月份叫做闰月就是很自然的了.

例如,1968年,农历为戊申年,这年八月以前各月中气所在的日期如下:

雨水	正月廿一
春分	二月廿二
谷雨	三月廿三
小满	四月廿五
夏至	五月廿六
大暑	六月廿八
处暑	七月三十
白露	闰七月十五
秋分	八月初二

这里的闰七月中只有一个节气白露,而没有中气.

下面列出从1949年到2020年农历闰月的情况:

1949	闰七月	1974	闰四月	1998	闰五月
1952	闰五月	1976	闰八月	2001	闰四月
1955	闰三月	1979	闰六月	2004	闰二月
1957	闰八月	1982	闰四月	2006	闰七月
1960	闰六月	1984	闰十月	2009	闰五月
1963	闰四月	1987	闰六月	2012	闰四月
1966	闰三月	1990	闰五月	2014	闰九月
1968	闰七月	1993	闰三月	2017	闰六月
1971	闰五月	1995	闰八月	2020	闰四月

3.2.6　日月食

前面已经介绍过朔望月,现在再介绍交点月.大家知道地球绕太阳转,月亮绕地球转.地球的轨道在一个平面上,称为黄道面.而月亮的轨道并不在这个平面上,因此月亮轨道和这黄道面有交点.具体地

说,月亮从地球轨道平面的这一侧穿到另一侧时有一个交点,再从另一侧穿回这一侧时又有一个交点,其中一个在地球轨道圈内,另一个在圈外,从圈内交点到圈内交点所需时间称为交点月.交点月约为27.2123天.当太阳、月亮和地球的中心在一直线上,这时就发生日食(图3-1)或月食(如果月亮在地球的另一侧).如图3-1,由于三点在一直线上,因此月亮一定在地球轨道平面上,也就是月亮在交点上;同时也是月亮全黑的时候,也就是朔.从这样的位置再回到同样的位置必需要有两个条件:从一交点到同一交点(这和交点月有关);从朔到朔(这和朔望月有关).现在我们来求朔望月和交叉点月的比.我们有

$$\frac{29.5306}{27.2123}=1+\frac{1}{11}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{4}+\frac{1}{2}+\frac{1}{9}+\frac{1}{1}+\frac{1}{25}+\frac{1}{2}.$$

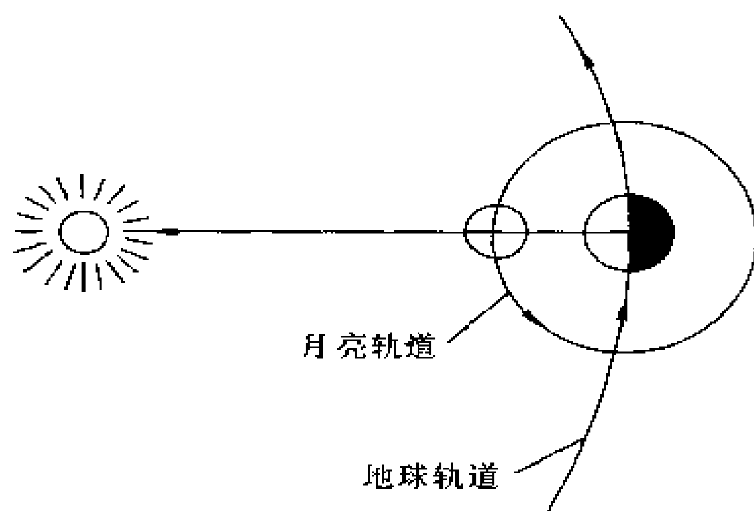


图3-1

考虑渐近分数

$$1+\frac{1}{11}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{4}=\frac{242}{223},$$

而 223×29.5306 天 = 6585 天 ≈ 18 年 11 天.这就是说,经过 242 个交点月或 223 个朔望月以后,太阳、月亮和地球又差不多回到了原来的相对位置.应当注意的是不一定这三个天体的中心准在一直线上时才出现日食或月食,稍偏一些也会发生,因此在这 18 年 11 天中会发生好多次日食和月食(约有 41 次日食和 29 次月食),虽然相邻两次日食(或月食)时间隔时间并不是一个固定的数,但是经过了 18 年 11 天以后,由于这三个天体又回到了原来的相对位置,因此在这 18 年 11 天中日食、月食发生的规律又重复实现了.这个交食(日食月食的总称)的周期称为沙罗周期."沙罗"就是重复的意思.求出了沙罗周期,就大大便于日食月食的测定.

3.2.7 日月合璧,五星联珠,七曜同宫

1962 年 2 月 5 日是春节.正在那天,天空中出现了一个非常罕见的现象,就是金、木、水、火、土五大行星在同一个方向上出现,而且就在这个方向上日食也正好发生.这种现象称为日月合璧,五星联珠,七曜同宫(图 3-2).这种奇异的现象几百年才会出现一次.

天文学家把"天"分为若干部分(这里的"天"指的是我们的视空间),每一部分称为一个星座.通过黄道面的共有 12 个星座,称为黄道 12 宫.1962 年 2 月 5 日这一天,金、木、水、火、土、日、月七个星球同时走到一个宫内.这就是七曜同宫.为什么称为五星联珠呢?因为在那一天人们从地球上看起来金、木、水、火、土五个行星的位置差不多在一起.但实际上有远有近,好象串成一串珠子一样.这种现象也称为五星聚.古人把视角不超过 45° 的情况就称为五星联珠.五星联珠被视为吉祥之兆.对这种现象远在两千多年前我国历史上就有了记载.在《汉书》律历志上是这样写的

复覆太初历,晦朔弦望皆最密,日月如合璧,五星如联珠.

还有一个注:

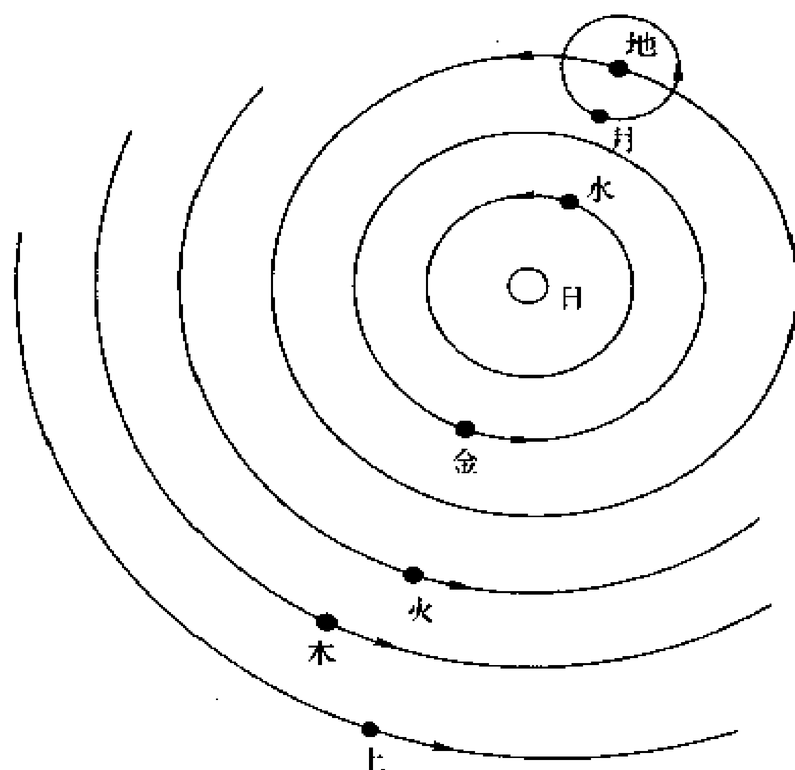


图 3－2

太初上元甲子夜半朔旦冬至时，七曜皆会聚斗牵牛分度，夜尽如合璧联珠也．

太初是汉武帝的年号，在公元前 104 年．这说明我们的祖先对天文现象早就作了精密的观察．

何时再发生联珠现象，也可用上面提供的方法算，当然复杂度增加了．

2.2.8 干支纪年

在我国的日历牌上通常有两部分：一是用阿拉伯数字表示的公历日期，另一是用汉文数字表示的农历日期．这两者之间常常用“农历壬子年三月小”，“农历丙午年二月大”等字样隔开，这里的“壬子”、“丙午”就叫做“干支”．查一查过去的历书就知道，“壬子”年对

应的是1972年，"丙午"年对应的是1966年．干支实际上是"天干"和"地支"的合称．甲、乙、丙、丁、戊、己、庚、辛、壬、癸十个字叫做"天干"；子、丑、寅、卯、辰、巳、午、未、申、酉、戌、亥十二个字叫做"地支"．把天干中的一个字摆在前面，后面配上地支中的一个字，就构成一对干支．如果天干以"甲"字开始，地支以"子"字开始组合，我们就可以得到六十对干支，这常叫做"六十干支"或"六十花甲子"．我们把六十干支表排列如下．

1	2	3	4	5	6	7	8	9	10
甲子	乙丑	丙寅	丁卯	戊辰	己巳	庚午	辛未	壬申	癸酉
11	12	13	14	15	16	17	18	19	20
甲戌	乙亥	丙子	丁丑	戊寅	己卯	庚辰	辛巳	壬午	癸未
21	22	23	24	25	26	27	28	29	30
甲申	乙酉	丙戌	丁亥	戊子	己丑	庚寅	辛卯	壬辰	癸巳
31	32	33	34	35	36	37	38	39	40
甲午	乙未	丙申	丁酉	戊戌	己亥	庚子	辛丑	壬寅	癸卯
41	42	43	44	45	46	47	48	49	50
甲辰	乙巳	丙午	丁未	戊申	己酉	庚戌	辛亥	壬子	癸丑
51	52	53	54	55	56	57	58	59	60
甲寅	乙卯	丙辰	丁巳	戊午	己未	庚申	辛酉	壬戌	癸亥

按照表的次序，每年用一对干支表示，这种纪年法叫"干支纪年法．"从古代文献来看，干支纪年至迟在东汉初期已经普遍使用，直到今天没有间断过．

干支纪年在我国历史学中广泛使用，特别是近代史中很多重要的历史事件的年代常用干支年表示．例如，甲午战争、庚子义和团起义、戊戌变法、辛亥革命等等．然而，在现代史中由于采取了公历纪年，干支纪年就不必要了．

把公历纪年换算成干支纪年，通常要查阅专门编制的年代对照表，这类书一般比较少，而查起来也很麻烦．下面介绍一个简单的计算公式，可以用来很容易的算出公历某年所对应的干支来．这个公式

是

$$n = x - 3 - 60m,$$

这里 n 是干支表中的序数，x 是所求年的公历纪年数，$m = 0,1,2,\cdots$，取整数值，适当选择 m 的值，使

$$0 < n \leqslant 60.$$

从得到的 n 就可立即从表中查出干支来.

例　求 1894 年的干支.

解　$x = 1894$，选取 $m = 31$，则

$$n = 1894 - 3 - 60 \times 31 = 31.$$

由干支表中查出，对应的干支是甲午. 1894 年正是甲午战争发生的年代.

这里有一点值得注意. 从公式的要求看，x 只能取公元 4 年以后的值，那么公元 4 年以前的干支怎么办？天文纪年法规定，公元元年记为 + 1 年，公元前一年记为 0 年，公元前 2 年记为 − 1 年，公元前 3 年记为 − 2 年，…. 把公元 4 年以前的 x 值按这个方法取值，而且 m 也可以取负整数，那么，公式(1) 仍旧成立.

例　求公元前 221 年的干支.

解　这时，依规定 $x = -220$，取 $m = -4$，则

$$n = (-220) - 3 - 60 \times (-4) = 17.$$

由干支表查出，干支为庚辰.

这是秦国完成统一，秦始皇称帝的那年.

反过来，从干支纪年换算公历纪年，要复杂一些. 因为干支经过 60 年就重复一次. 同一干支对应一系列公历纪年，它们之间相差 60 的整数倍，为了解决这种不确定性，需要参考其它历史因素.

§3.3　连分数的性质

3.3.1　渐近分数的性质

前面以连分数为工具研究了许多天文现象，简单而漂亮地得出许多有用的结果，同时我们还注意到连分数还有一些很好的性质．现在我们就着手对连分数的理论作些初步探讨．先从 π 的渐近分数开始，看它能给我们些什么启示．

$\pi = 3.1415926\cdots$，它的渐近分数是

$$3, \frac{22}{7} = 3.1428, \frac{333}{106} = 3.1415094, \frac{355}{113} = 3.1415929.$$

与 π 的准确值相比较，我们看到，

$$3 < \frac{333}{106} < \pi < \frac{355}{113} < \frac{22}{7}.$$

如果将各阶渐近分数分别记为

$$\frac{p_0}{q_0}, \frac{p_1}{q_1}, \cdots, \frac{p_n}{q_n}, \cdots,$$

那么，它们似乎呈现下述规律：

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \pi < \cdots < \frac{p_{2n+1}}{q_{2n+1}} < \cdots < \frac{p_1}{q_1}.$$

更清楚些，可写为：

1) $\frac{p_{2n}}{q_{2n}} < \pi$ 是递增序列；$\frac{p_{2n+1}}{q_{2n+1}} > \pi$ 是递减序列；

2) $\frac{p_{2n}}{q_{2n}} \to \pi, \frac{p_{2n+1}}{q_{2n+1}} \to \pi (n \to \infty)$.

前面曾指出，在所有分母小于 8 的分数中以 22/7 最接近于 π．这样，渐近分数还有性质

3) $\frac{p_n}{q_n}$ 是对 π 的最佳逼近，即在所有分母 $q \leqslant q_n$ 的分数中，$\frac{p_n}{q_n}$ 和

π 最接近.

下面我们证明这些性质对任何连分数都成立，但要花一点力气.

研究渐近分数，首先看看它有无规律，能不能找出渐近分数的表达式. 答案是肯定的，但找到这个答案花费了数学家不少的心思.

3.3.2 渐近分数的表达式

设 $\alpha > 0$ 是一个实数，其连分数展式为

$$\alpha = a_0 + \frac{1}{a_1 +}\,\frac{1}{a_2 +}\cdots,$$

它的前三个渐近分数是

$$\frac{a_0}{1},\ \frac{a_1 a_0 + 1}{a_1},\ \frac{a_2(a_1 a_0 + 1) + a_0}{a_2 a_1 + 1}.$$

第一个是明显的，第二个请读者自己算，我们算一算第三个.

$$a_0 + \frac{1}{a_1 +}\,\frac{1}{a_2} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}}$$

$$= a_0 + \cfrac{1}{\cfrac{a_1 a_2 + 1}{a_2}} = a_0 + \frac{a_2}{a_1 a_2 + 1} = \frac{a_0 a_1 a_2 + a_0 + a_2}{a_1 a_2 + 1}$$

$$= \frac{a_2(a_0 a_1 + 1) + a_0}{a_1 a_2 + 1}.$$

在一般情况下，我们有

定理 1 若命

$$p_0 = a_0,\ p_1 = a_1 a_0 + 1,\ p_n = a_n p_{n-1} + p_{n-2} (n > 1), \quad (1)$$

$$q_0 = 1,\ q_1 = a_1,\ q_n = a_n q_{n-1} + q_{n-2} (n > 1), \quad (2)$$

则 p_n / q_n 就是 α 的第 n 个渐近分数.

证 用归纳法证明. 当 $n = 2$ 时，

$$\frac{p_2}{q_2} = \frac{a_2 p_1 + p_0}{a_2 q_1 + q_0} = \frac{a_2(a_1 a_0 + 1) + a_0}{a_2 a_1 + 1},$$

与上面的公式一致.设 $n=k$ 时公式成立,今证当 $n=k+1$ 时公式也成立.a 的第 k 个渐近分数是

$$a_0+\frac{1}{a_1}+\cdots+\frac{1}{a_k}.$$

a 的第 $k+1$ 个渐近分数是

$$a_0+\frac{1}{a_1}+\cdots+\frac{1}{a_k}+\frac{1}{a_{k+1}}.$$

两者的差别仅在于将 a_k 换成 $a_k+\frac{1}{a_{k+1}}$.如果 a 第 k 个渐近分数是

$$\frac{p_k}{q_k}=\frac{a_kp_{k-1}+p_{k-2}}{a_kq_{k-1}+q_{k-2}},$$

则第 $k+1$ 渐近分数是

$$\begin{aligned}\frac{p_{k+1}}{q_{k+1}}&=\frac{\left(a_k+\frac{1}{a_{k+1}}\right)p_{k-1}+p_{k-2}}{\left(a_k+\frac{1}{a_{k+1}}\right)q_{k-1}+q_{k-2}}\\&=\frac{(a_ka_{k+1}+1)p_{k-1}+a_{k+1}p_{k-2}}{(a_ka_{k+1}+1)q_{k-1}+a_{k+1}q_{k-2}}\\&=\frac{a_ka_{k+1}p_{k-1}+a_{k+1}p_{k-2}+p_{k-1}}{a_ka_{k+1}q_{k-1}+a_{k+1}q_{k-2}+q_{k-1}}\\&=\frac{a_{k+1}(a_kp_{k-1}+p_{k-2})+p_{k-1}}{a_{k+1}(a_kq_{k-1}+q_{k-2})+q_{k-1}}\\&=\frac{a_{k+1}p_k+p_{k-1}}{a_{k+1}q_k+q_{k-1}}.\end{aligned}$$

定理证毕.

有了渐近分数的表达式后,就可以研究渐近分数之间有什么联系了.找到它们之间的联系就可以比较它们的大小,从而增减的信息也就知道了.下面的定理告诉我们,相邻渐近分数之间有非常简单而明确的关系.性质本身十分有趣.

定理 2 $p_nq_{n-1} - q_np_{n-1} = (-1)^{n-1} \quad (n > 0)$, (3)

$p_nq_{n-2} - q_np_{n-2} = (-1)^na_n \quad (n > 1)$. (4)

证 用归纳法证明.

证(3) 当 $n = 1$ 时, 公式(3) 成立:

$$p_1q_0 - q_1p_0 = a_1a_0 + 1 - a_1a_0 = 1.$$

若 $n = k$ 时(3) 成立,则由(1),(2),

$$\begin{aligned} & p_{k+1}q_k - q_{k+1}p_k \\ &= (a_{k+1}p_k + p_{k-1})q_k - (a_{k+1}q_k + q_{k-1})p_k \\ &= p_{k-1}q_k - q_{k-1}p_k \\ &= -(p_kq_{k-1} - q_kp_{k-1}) \\ &= -(-1)^{k-1} = (-1)^k. \end{aligned}$$

证(4) 今用(3) 证(4).

$$\begin{aligned} & p_nq_{n-2} - q_np_{n-2} \\ &= (a_np_{n-1} + p_{n-2})q_{n-2} - (a_nq_{n-1} + q_{n-2})p_{n-2} \\ &= a_np_{n-1}q_{n-2} - a_nq_{n-1}p_{n-2} \\ &= a_n(p_{n-1}q_{n-2} - q_{n-1}p_{n-2}) \\ &= (-1)^na_n. \end{aligned}$$

证毕.

系 1 $(p_n, q_n) = 1$.

证 由(3), p_n, q_n 的公因数一定整除$(-1)^{n-1}$,它一定是 1 或 (-1).

系 1 指出,渐近分数都是既约分数.

(3) 的两边除以 $q_{n-1}q_n$,(4) 的两边除以 $q_{n-2}q_n$,可得

系 2

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_nq_{n-1}}, \tag{5}$$

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^na_n}{q_nq_{n-2}}. \tag{6}$$

有了系 2 就可以比较两个相邻的渐近分数，或两个相隔的渐近分数的大小了.

3.3.3　渐近分数的极限

设 $a > 0$，于是 $a_0 \geqslant 0, a_n \geqslant 1 (n = 1, 2, \cdots)$. 令

$$\alpha_n = a_n + \frac{1}{a_{n+1}} + \frac{1}{a_{n+2}} + \cdots. \tag{7}$$

由

$$a = a_0 + \frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n} + \frac{1}{a_{n+1}} + \cdots \tag{8}$$

和

$$\frac{p_n}{q_n} = a_0 + \frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n},$$

可看出，a 与它的第 n 个渐近分数的差别仅在于将 p_n/q_n 中的 a_n 换成 α_n. 这样一来，由定理 1，立刻得到

定理 3

$$a = \alpha_0, a = \frac{\alpha_1 a_0 + 1}{\alpha_1}, a = \frac{\alpha_n p_{n-1} + p_{n-2}}{\alpha_n q_{n-1} + q_{n-2}} \quad (n \geqslant 2). \tag{9}$$

这是个预备性的定理. 我们的目的在于找出 a 与它的第 n 个渐近分数的差，有了差就可以考虑极限问题了.

定理 4

$$a - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})} = \frac{(-1)^n \alpha_{n+2}}{q_n(\alpha_{n+2}q_{n+1} + q_n)}. \tag{10}$$

注　(10) 式实际上是两个公式，需要分别证明.

证　今以两种方式考察 a 与渐近分数的差. 由(9)，a 可有两种表示：

$$a = \frac{\alpha_n p_{n-1} + p_{n-2}}{\alpha_n q_{n-1} + q_{n-2}}, a = \frac{\alpha_{n+1} p_n + p_{n-1}}{\alpha_{n+1} q_n + q_{n-1}}.$$

用这两个表达式分别与 $\frac{p_n}{q_n}$ 作差，就可得到(10). 下面的任务只是计算.

$$a-\frac{p_n}{q_n}=\frac{\alpha_{n+1}p_n+p_{n-1}}{\alpha_{n+1}q_n+q_{n-1}}-\frac{p_n}{q_n}$$

$$=\frac{\alpha_{n+1}p_nq_n+p_{n-1}q_n-\alpha_{n+1}p_nq_n-p_nq_{n-1}}{q_n(\alpha_{n+1}q_n+q_{n-1})}$$

$$=\frac{-(p_nq_{n-1}-q_np_{n-1})}{q_n(\alpha_{n+1}q_n+q_{n-1})}=\frac{(-1)^n}{q_n(\alpha_{n+1}q_n+q_{n-1})}.$$

又，

$$a-\frac{p_n}{q_n}=\frac{\alpha_{n+2}p_{n+1}+p_n}{\alpha_{n+2}q_{n+1}+q_n}-\frac{p_n}{q_n}$$

$$=\frac{\alpha_{n+2}p_{n+1}q_n+p_nq_n-\alpha_{n+2}p_nq_{n+1}-p_nq_n}{q_n(\alpha_{n+2}q_{n+1}+q_n)}$$

$$=\frac{\alpha_{n+2}(p_{n+1}q_n-q_{n+1}p_n)}{q_n(\alpha_{n+2}q_{n+1}+q_n)}=\frac{(-1)^n\alpha_{n+2}}{q_n(\alpha_{n+2}q_{n+1}+q_n)}.$$

证毕.

由定理 4，当 $n=2m+1$ 时，

$$a-\frac{p_{2m+1}}{q_{2m+1}}=-\frac{1}{q_n(\alpha_{n+1}q_n+q_{n-1})}<0\ \Rightarrow\ a<\frac{p_{2m+1}}{q_{2m+1}}.\quad(11)$$

当 $n=2m$ 时，$(-1)^{2m}=1$，从而

$$a-\frac{p_{2m}}{q_{2m}}>0\ \Rightarrow\ \frac{p_{2m}}{q_{2m}}<a.\quad(12)$$

利用(6)，当 $n=2m+1$ 时，

$$\frac{p_{2m+1}}{q_{2m+1}}-\frac{p_{2m-1}}{q_{2m-1}}=\frac{-a_n}{q_{2m+1}q_{2m-1}}<0\ \Leftrightarrow\ \frac{p_{2m+1}}{q_{2m+1}}<\frac{p_{2m-1}}{q_{2m-1}}.$$

与(11)结合起来就得出

$$\frac{p_1}{q_1}>\frac{p_3}{q_3}>\cdots>\frac{p_{2n+1}}{q_{2n+1}}>\cdots>a.\quad(13)$$

由(6)，当 $n=2m$ 时，

$$\frac{p_{2m}}{q_{2m}}-\frac{p_{2m-2}}{q_{2m-2}}=\frac{a_n}{q_{2m}q_{2m-2}}>0\ \Rightarrow\ \frac{p_{2m}}{q_{2m}}>\frac{p_{2m-2}}{q_{2m-2}}.$$

与(12)结合起来就得出

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \alpha. \tag{14}$$

(13),(14)结合起来,有

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \alpha$$

$$< \cdots < \frac{p_{2n+1}}{q_{2n+1}} < \cdots < \frac{p_3}{q_3} < \frac{p_1}{q_1}. \tag{15}$$

由(5)还可得到

$$\left|\frac{p_{2n}}{q_{2n}} - \frac{p_{2n-1}}{q_{2n-1}}\right| = \frac{1}{q_{2n}q_{2n-1}}. \tag{16}$$

定理 5 $q_n \geqslant n$.

证 由定理1,

$$q_0 = 1, q_1 = a_1 \geqslant 1, a_n \geqslant 1,$$

$$q_n = a_n q_{n-1} + q_{n-2} \geqslant q_{n-1} + 1.$$

由此,利用数学归纳法可证,$q_n \geqslant n$.

定理5指出,$q_n \to \infty (n \to \infty)$.

定理 6 $\alpha = \lim_{n \to \infty} \frac{p_n}{q_n}$.

证 由定理4,

$$\left|\alpha - \frac{p_n}{q_n}\right| = \left|\frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})}\right| \leqslant \frac{1}{q_n^2} \to 0.$$

3.3.4 连分数的几何解释

一个无理数的连分数的渐近分数收敛到这个无理数,这一事实有一个有趣的几何解释.这是1897年德国数学家F.克莱因给出的.克莱因不仅是一个卓越的数学家,而且善于写科普著作.他的一些著作至今还在重印和使用.

设 $\alpha > 0$ 是一个无理数,它的第 n 个渐近分数是 p_n/q_n.在坐标

纸上标出横坐标 x 和纵坐标 y 是整数的点 (x, y)，这些点称为格点．想象在这些点处插上一个大头针．然后画一条直线 $y = \alpha x$．这条直线不通过任何一个格点；因为，如果有一个坐标为整数的点满足方程，就会推出 α 是有理数，但我们已假定 α 为无理数．现在想象，把一条细线的一端绑在直线 $y = \alpha x$ 的无穷遥远的一点上，而把另一端拿在手中．把线拉直，并使手中的一点位于原点．从原点向上移动手，这条线就会从上面碰到一些大头针；向下移动手，这条线就会从下面碰到另一些大头针，如图 3－3 所示．

从上面触及这条线的大头针具有坐标 (q_1, p_1)，(q_3, p_3)，(q_5, p_5)，…．它们分别对应于奇次渐近分数

$$\frac{p_1}{q_1}, \frac{p_3}{q_3}, \frac{p_5}{q_5}, \cdots.$$

从下面触及这条线的大头针具有坐标 (q_0, p_0)，(q_2, p_2)，(q_4, p_4)，… 它们分别对应于偶次渐近分数

$$\frac{p_0}{q_0}, \frac{p_2}{q_2}, \frac{p_4}{q_4}, \cdots.$$

用折线 L_1 把 (q_1, p_1)，(q_3, p_3)，(q_5, p_5)，… 连接起来，用折线 L_2 把 (q_0, p_0)，(q_2, p_2)，(q_4, p_4)… 连接起来．这两条折线向外走得越远，它们就越逼近直线 $y = \alpha x$．

例 图 3－3 画出了数

$$\alpha = \frac{1+\sqrt{5}}{2} = 1 + \frac{1}{1} + \frac{1}{1} + \cdots$$

的连分数展式的克莱因图．

习题 构造 $(\sqrt{5} - 1)/2$ 的连分数展式的克莱因图．

3.3.5 最佳逼近

连分数提供了对实数的最佳逼近，因而十分有用．

定理 7 设 α 是任一实数，p_k/q_k 是 α 的第 k 个渐近分数，那么在分母 $q \leqslant q_k$ 的一切有理数中，p_k/q_k 是 α 的最好有理近似值，即若

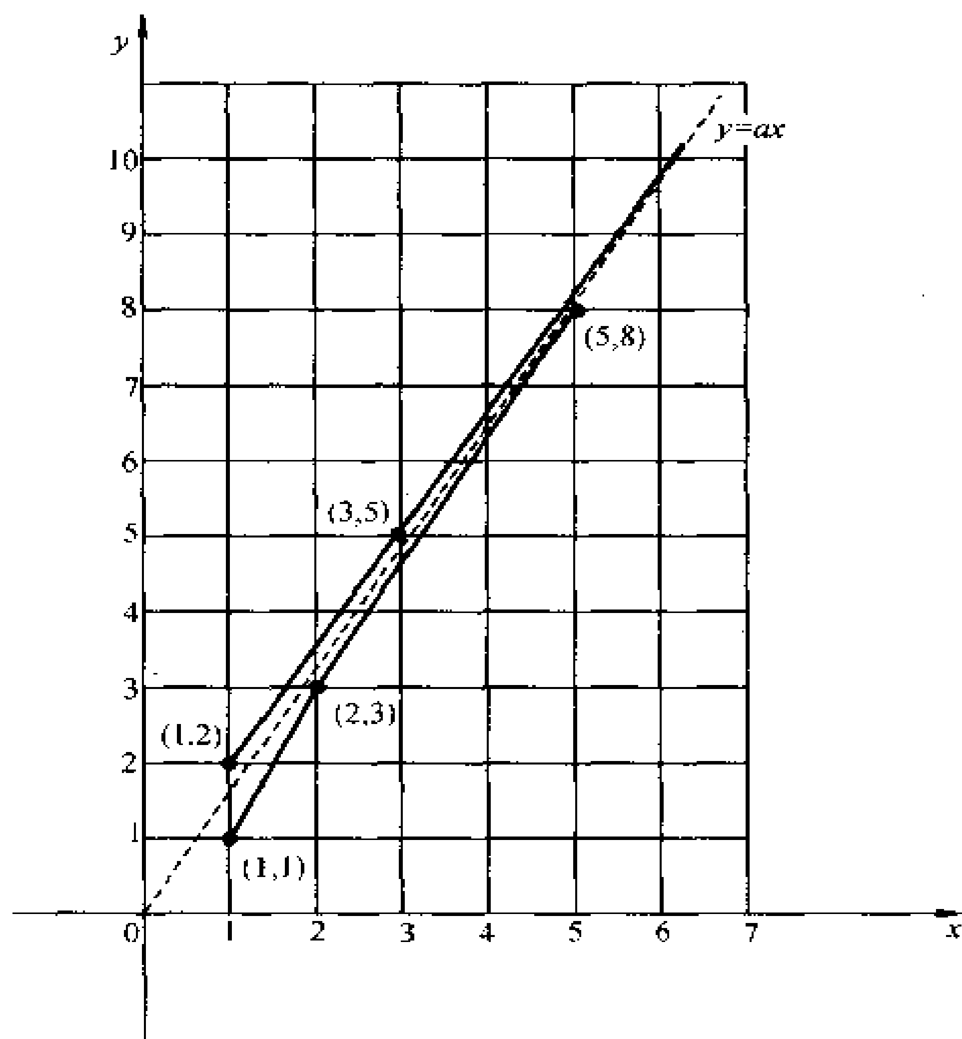


图 3 − 3

p/q 是一个满足 $0 < q \leqslant q_k$ 的有理数，则

$$\left| a - \frac{p_k}{q_k} \right| \leqslant \left| a - \frac{p}{q} \right|. \tag{17}$$

证　先讲一讲证明的基本思想. 若 $a = p_k/q_k$，则定理已经成立（我们只假定 a 是实数. 所以它可以是有理数，也可以是无理数. 当 a 是有理数的时候，就可能出现 $a = p_k/q_k$ 的情况）. 因此只需考虑 a

$\neq p_k/q_k$ 的情况. 这时 α 有第 $k+1$ 个渐近分数. 不妨设 $p_k/q_k < p_{k+1}/q_{k+1}$($p_{k+1}/q_{k+1} < p_k/q_k$ 的情况可完全类似地讨论). 前面已经证明,

$$\frac{p_k}{q_k} < \alpha \leqslant \frac{p_{k+1}}{q_{k+1}}$$

如图 3－4 所示, α 落在区间 $\left(\frac{p_k}{q_k}, \frac{p_{k+1}}{q_{k+1}}\right]$ 内. 证明可以分为两步: 1) p/q 落在区间的外面, 或者落在左边(图 3－5), 或者落在右边(图 3－6); 2) 若 p/q 落在区间的左边, 则

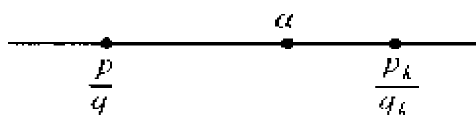


图 3－4

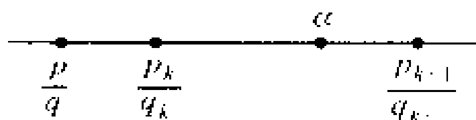


图 3－5

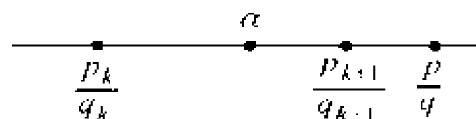


图 3－6

$$\frac{p}{q} \leqslant \frac{p_k}{q_k} \Rightarrow \alpha - \frac{p_k}{q_k} \leqslant \alpha - \frac{p}{q}.$$

当 p/q 落在区间的右边时, 后面证明.

1) 先证明: 若 $0 < q \leqslant q_k$, 则

$$\frac{p}{q} \leqslant \frac{p_k}{q_k} \quad \text{或} \quad \frac{p_{k+1}}{q_{k+1}} < \frac{p}{q}. \tag{18}$$

今用反证法证之. 如果(18)不成立,则有

$$\frac{p_k}{q_k} < \frac{p}{q} \leqslant \frac{p_{k+1}}{q_{k+1}}.$$

由于 p_{k+1}/q_{k+1} 是既约分数,而 $q \leqslant q_k < q_{k+1}$,所以上式中等式不能成立. 于是

$$\frac{p_k}{q_k} < \frac{p}{q} < \frac{p_{k+1}}{q_{k+1}}.$$

因此,

$$\frac{p}{q} - \frac{p_k}{q_k} = \frac{pq_k - qp_k}{qq_k} \geqslant \frac{1}{qq_k},$$

$$\frac{p_{k+1}}{q_{k+1}} - \frac{p}{q} = \frac{qp_{k+1} - pq_{k+1}}{qq_{k+1}} \geqslant \frac{1}{qq_{k+1}},$$

又,由假设 $q_{k+1} + q_k > 2q$,从而

$$\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} = \frac{p_{k+1}}{q_{k+1}} - \frac{p}{q} + \frac{p}{q} - \frac{p_k}{q_k} > \frac{1}{qq_k} + \frac{1}{qq_{k+1}}$$

$$= \frac{q_{k+1} + q_k}{qq_kq_{k+1}} > \frac{2q}{qq_kq_{k+1}} = \frac{2}{q_kq_{k+1}}.$$

与定理 2 的系 2 相矛盾. 这就证明了(18).

2) 若

$$\frac{p_{k+1}}{q_{k+1}} < \frac{p}{q},$$

则

$$\left|\alpha - \frac{p}{q}\right| \geqslant \left|\frac{p_{k+1}}{q_{k+1}} - \frac{p}{q}\right| = \left|\frac{qp_{k+1} - pq_{k+1}}{qq_{k+1}}\right| \geqslant \frac{1}{qq_{k+1}} \geqslant \frac{1}{q_kq_{k+1}}.$$

再算 α 与 $\frac{p_k}{q_k}$ 的距离. 由定理 4,

$$\left|a-\frac{p_k}{q_k}\right|=\frac{1}{q_k(\alpha_{k+1}q_k+q_{k-1})}.$$

但 $\alpha_{k+1}>a_{k-1}$ 所以 $\alpha_{k+1}q_k+q_{k-1}>a_{k-1}q_k+q_{k-1}=q_{k+1}$. 从而

$$\left|a-\frac{p_k}{q_k}\right|=\frac{1}{q_k(\alpha_{k+1}q_k+q_{k-1})}<\frac{1}{q_kq_{k+1}}.$$

所以
$$\left|a-\frac{p_k}{q_k}\right|<\left|a-\frac{p}{q}\right|.$$

证毕.

3.3.6 方程 $x^2=ax+1$ 的解

连分数可以用来求任何代数方程的正根的近似解,条件是该方程有这样的解.我们来考察二次代数方程

$$x^2=ax+1. \tag{19}$$

若 $a>0$,则这个方程的正根具有连分数展式

$$x=a+\frac{1}{a}+\frac{1}{a}+\cdots.$$

为了看出这一点,我们只要在(19)两边除以 x 就得到

$$x=a+\frac{1}{x}.$$

所以
$$x=a+\frac{1}{a+\frac{1}{x}}=a+\frac{1}{a}+\frac{1}{a}+\cdots.$$

例如,当 $a=1$ 时,方程 $x^2=x+1$ 有正根

$$\alpha=\frac{1+\sqrt{5}}{2}=1+\frac{1}{1}+\frac{1}{1}+\cdots. \tag{20}$$

数 α 是产生黄金分割的数,而黄金分割是构造正五边形的关键.下面讲正多边形的作图时,还会遇到它.

3.3.7 斐波那契级数

数学的各个领域常常奇妙而出乎意料地联系在一起.(20)式中的连分数是最简单的无限连分数,它的渐近分数是

$$\frac{1}{1},\frac{2}{1},\frac{3}{2},\frac{5}{3},\frac{8}{5},\frac{13}{8},\cdots. \tag{21}$$

它们的分子和分母由下述整数序列组成：

$$1,1,2,3,5,8,13,21,34,\cdots. \tag{22}$$

在这些数中，从第三项开始，每一个数都是它前面的两个数的和；例如，$2=1+1$，$8=5+3$，等等．这些数以斐波那契级数而著称．

斐波那契级数出现在意大利数学家斐波那契(Fibonacci；1174－1250)在1202年所著的《算盘书》中．书中是这样提出问题的："如果每对兔子每月能繁殖一对子兔，而子兔在出生后第二个月就有生殖能力，第三个月就生产一对兔子，以后每个月生产一对，假定每对兔子都是一雌一雄．试问一对兔子一年能繁殖多少对兔子？"由这个问题得出的序列就是上面的(22)．出人意料的是，这个序列在许多场合都出现．因此，我们需要对它作些探讨．

数列(22)中的每一个数叫做斐波那契数．若第 n 个斐波那契数记为 F_n，则我们有

$$F_0=1,F_1=1,F_2=2,F_3=3,F_4=5,\cdots.$$

这个序列有下面的递推关系

$$F_{n+2}=F_{n+1}+F_n\quad(n=0,1,2,\cdots). \tag{23}$$

由此出发借助数学归纳法可推得通项公式

$$F_n=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]. \tag{24}$$

这个公式是法国数学家比内(Binet)求出的．

证　1) $n=0$．这时

$$\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$$

$$=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{1}-\left(\frac{1-\sqrt{5}}{2}\right)^{1}\right]$$

$$= \frac{\sqrt{5}}{\sqrt{5}} = 1 = F_0.$$

2）为了证明下面的情况，我们需要计算

$$\left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = 1 + \frac{1+\sqrt{5}}{2}. \tag{25}$$

类似地，

$$\left(\frac{1-\sqrt{5}}{2}\right)^2 = 1 + \frac{1-\sqrt{5}}{2}, \tag{26}$$

从而

$$\left(\frac{1+\sqrt{5}}{2}\right)^2 - \left(\frac{1-\sqrt{5}}{2}\right)^2 = \sqrt{5}.$$

3）$n = 1$. 这时

$$\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$$

$$= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{2} - \left(\frac{1-\sqrt{5}}{2}\right)^{2}\right]$$

$$= \frac{\sqrt{5}}{\sqrt{5}} = 1 = F_1.$$

4）设 n 是任意自然数，并假定公式(24)对一切 $k < n$ 成立. 将归纳法假设应用于 n，我们得到，

$$\begin{aligned} F_n &= F_{n-2} + F_{n-1} \\ &= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\right] \\ &\quad + \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n} - \left(\frac{1-\sqrt{5}}{2}\right)^{n}\right] \\ &= \frac{1}{\sqrt{5}}\left\{\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} + \left(\frac{1+\sqrt{5}}{2}\right)^{n}\right]\right. \\ &\quad \left. - \left[\left(\frac{1-\sqrt{5}}{2}\right)^{n-1} + \left(\frac{1-\sqrt{5}}{2}\right)^{n}\right]\right\} \end{aligned}$$

$$= \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^{n-1}\left[1+\frac{1+\sqrt{5}}{2}\right]\right.$$

$$\left. - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\left[1+\frac{1-\sqrt{5}}{2}\right]\right\}.$$

由(25),(26),

$$= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-1}\left(\frac{1+\sqrt{5}}{2}\right)^{2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\left(\frac{1-\sqrt{5}}{2}\right)^{2}\right]$$

$$= \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right].$$

公式(24)得证.

斐波那契数是大自然的一个基本模式,它出现在许多场合.在花的花瓣中存在斐波那契模式.几乎所有的花,其花瓣数都是斐波那契数.例如百合花的花瓣有3瓣;毛茛属的植物有5瓣花;许多翠雀属植物有8瓣花;万寿菊的花有13瓣;紫菀属的植物有21瓣花;大多数雏菊有34,55,89瓣花.在向日葵的花盘内葵花子的螺旋模式中也可以发现斐波那契级数.虽然这一模式几个世纪前已被注意到,此后又被广泛研究,但真正满意的解释直到1993年才给出.

斐波那契数有许多应用.它可应用于分割问题.例如,把一个正方形分割为若干个不相重的正方形.在辗转相除法中,为了求出两个自然数的最大公约数,需要进行一系列除法.我们需要知道除法次数的上限.对此,G.拉梅给出一个巧妙的定理:为了求出两个自然数的最大公约数,所需要进行的除法的次数绝不大于较小自然数的位数的五倍.这个定理的证明首先要用到斐波那契级数的某些性质.斐波那契级数还有许多其它应用,以致1963年成立了斐波那契协会,出版了《斐波那契季刊》

第四章　素数定理与哥德巴赫猜想

算术给予我们一个用之不尽的,充满有趣真理的宝库.这些真理不是孤立的,而是以相互最密切的关系并立着,而且随着科学的每一成功的进展,我们不断地发现这些真理间的新的,完全意外的接触点.

G.F.高斯

数学,科学的皇后;数论,数学的皇后.

G.F.高斯

§4.1　初等数论初步

4.1.1　数论是什么

数论是研究整数性质的一个数学分支.它主要包括初等数论,解析数论,代数数论,丢番图逼近,超越数论等,还有其它分支.现代数论已经深入到数学的一切分支.

初等数论以算术方法为主要方法.初等数论中某些问题的研究促成新的数学分支的产生.如对不定方程和高次互反定律的研究促进了代数数论和类域论的形成和发展.

初等数论中仍有许多问题没有解决.

近几十年来初等数论在计算机科学,组合数学,代数编码、计算方法,信号的数字处理等领域内得到广泛的应用.

解析数论以分析工具为其主要工具,涉及到分析与数论的交互作用.所研究的问题分为三类:

1) 用分析方法证明整数的一些定性性质.最著名的例子是哥德

巴赫猜想.

2）素数分布.

3）用整数性质分析解析的性质.例如在有理数域和代数数域上的模函数等.

代数数论.普通数论研究整数,代数数论研究更广泛的一类数.

我们将选择几个最著名的问题进行讨论,它们分别来自数论的不同分支.

4.1.2 数论的一个特点:表面简单,实际难

数论表面上简单,它的主要定理可以表述得人人容易理解,但证明起来却需要极其艰深和复杂的数学工具.这里指出,研究经典数论问题必需有坚实的数学基础,否则会劳而无功.请读者切记此言.

4.1.3 素数与合数

自然数的一个最重要的性质是,一些数能分解为两个或多个较小的数的乘积.

例如

$$6 = 2 \cdot 3,\ 30 = 2 \cdot 15 = 3 \cdot 10 = 2 \cdot 3 \cdot 5.$$

另一些则不能,如 3,5,7,11 等.当

$$c = a \cdot b,$$

时,a 或 b 就称为 c 的因数或除数,记为 $a|c$,$b|c$.若 a 不能整除 c,记为 $a \nmid c$.

易见,对于任何自然数 a 都有 $1|a$, $a|0$, $a|a$.这说明 a 至少有因数 1 和 a.

基于上面的观察,我们将自然数分为三类:

1）1;

2）p,只有自然数 1 和 p 是它的因数;

3）n,有两个以上大于 1 的因数.

2）类中的数叫素数,又叫质数.如 2,3,5,7,11,13,17,⋯.3）类

中的数叫合数，如 4,6,8,9,24,56,65,….

4.1.4 素数表

素数在数论中起着中心的作用．下面将讲述的算术基本定理指出，任何大于 1 的整数，或者本身就是素数，或者可以以唯一的方式写成素数的乘积．所以，把素数看成建筑整数大厦的砖石是再恰当不过了．在这个意义上，数论中的素数犹如化学中的原子，都是同样值得认真研究的．由于素数的中心作用，造一张素数表，给出不超过一个给定自然数 N 的所有素数就是一件很有意义的事情了．为此，先引进一条引理

引理 每一个合数 n 至少有一个素因数 $p \leqslant \sqrt{n}$．

证 设 p 是 n 的最小素因数，并设 $n = pn_1$，这里 $n_1 > p$．如果 $p > \sqrt{n}$，则 $p \cdot n_1 > \sqrt{n} \cdot \sqrt{n} = n$．这个矛盾指出了引理成立．

有了引理我们就可以造素数表了．任给一个正整数 N，可以按照下述方法求出一切不超过 N 的素数：把不超过 N 的一切正整数按大小关系排成一串

$$1,2,3,4,\cdots,N.$$

首先划去 1，第一个留下的是 2，它是一个素数：

~~1~~，2，3，4，5，6，7，8，9，10，…，N.

其次，从 2 起隔一位划去一数，这样就划去了 2 的一切倍数，

~~1~~，2，3，~~4~~，5，~~6~~，7，~~8~~，9，~~10~~，…，N.

第一个留下未划去的是 3，它不是 2 的倍数，因此是一个素数．然后从 3 起隔两位划去一数，所划去的数就是 $3+3m$（$m = 1,2,\cdots$），它们是 3 的一切倍数（3 本身除外）：

~~1~~，2，3，~~4~~，5，~~6~~，7，~~8~~，~~9~~，~~10~~，…，N.

第一个留下未划去的是 5，它不是小于它的素数（2 及 3）的倍数，因此它是素数．然后从 5 起每隔 $5-1=4$ 位划去一数，所划去的数是 5

$+5m(m=1,2,\cdots)$,也就是 5 的一切倍数(5 本身除外):

~~1~~, 2 , 3 , ~~4~~, 5 , ~~6~~, 7 , ~~8~~, ~~9~~, ~~10~~, 11, …, N.

如此继续进行,所划去的都是合数,第一个留下的都不是比它小的素数的倍数,因此总是一个素数.用这种方法可以逐一的把素数求出来.这种方法好像用筛子筛出素数一样,称为埃拉多斯染尼(Eratosthenes)筛法.

要求出不超过 N 的一切素数,根据引理,只需把不超过 $\sqrt{N}$ 的素数的倍数划去就行了,这因为不超过 N 的合数的最小素因数总是不超过 $\sqrt{N}$ 的.

为了更清楚的了解素数表的造法,我们举 $N=60$ 为例,造出小于 60 的素数表.

例　求不超过 60 的全体素数.

解　因为不超过 $\sqrt{60}<8$ 的素数是 2,3,5,7,所以在 2,3,…,60 中,留下 2,3,5,7,依次划去 2,3,5,7 的倍数:

~~1~~ 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~
19 ~~20~~ ~~21~~ ~~22~~ 23 24 ~~25~~ 26 27 ~~28~~ 29 ~~30~~ 31 ~~32~~ ~~33~~
34 ~~35~~ 36 37 ~~38~~ ~~39~~ ~~40~~ 41 ~~42~~ 43 ~~44~~ ~~45~~ ~~46~~ 47 ~~48~~
~~49~~ 50 ~~51~~ ~~52~~ 53 ~~54~~ ~~55~~ 56 ~~57~~ ~~58~~ 59 ~~60~~

留下的数

2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,59.

就是不超过 60 的全体素数.

上面所讲的就是著名的埃拉多斯染尼筛法.早在公元前 300 年左右,埃氏就提出这一筛法.素数表就是根据这一方法略加改变而造出来的.埃氏筛法的改进与发展是近代解析数论的重要工具之一.

这里自然出现的一个问题是,素数个数是有限的还是无限的?如果是有限的,就可以造一张素数表,囊括一切素数,使用起来就方便了.但是素数个数是无限的.大约在 2300 年前欧几里得就证明了存在无限多个素数.尽管如此,迄今为止还没有发现素数的模型或产生

素数的有效公式.因而寻找大的素数必须借助计算机一个一个地找.寻求大素数是数论研究的重要课题之一.

读者可能会产生一个疑问:找大素数有什么用?甚至会问研究数论有什么用?这的确是一个很有意义的问题,值得问.在二次世界大战以前,许多数学家都认为纯数学,特别像数论这样的学科是无用的.如英国著名数学家哈代就说过,数学是一种"完全清白而无害的职业".他还说:"真正的数学对战争毫无影响,至今还没有人能发现有什么火药味的东西是数论或相对论造成的,而且很多年也不会有人发现这类事情."

"文化大革命"前,对数论的这种认识在我国也严重地存在着.由于认为数论是理论脱离实践的,所以那时数论的教学和研究受到歧视.

二次大战以后,数学应用的面貌发生了根本性的变化.数学已经渗透到包括社会科学在内的社会的一切领域中去了.1945 年原子弹的蘑菇云使人们,也使哈代本人生前看到了相对论不可能与战争有关的预言的可怕的否定.他所钟爱的数论在 1982 年以来也已成为应用于能控制成千上万颗核导弹的密码系统的理论基础.

现在最好的密码是用素数制造的,极难破译.这也不难理解.用计算机算两个 100 位的素数的乘积是一件很容易的事情.但如果给定一个 200 位的数,让你找出它是哪两个素数的乘积,那就困难多了.数学家至今还没有发现一种有效的方法去迅速分解一个合数.所以用素数作密钥,安全度非常高.特别是,如果你知道一个别人不知道的大素数,你用这个素数作密钥,别人就不可能破译你的密码.

4.1.5 算术基本定理

一个合数 c 可以表示为乘积 $c = a \cdot b$,其中 a,b 两个数皆不为 1.a,b 中还可能有合数.若 a 是合数,则可以进一步分解:$a = a_1 \cdot a_2$.对 b 也一样.这样我们继续分解,直到不能分解为止.这种

分解一定会到某一步停止，因为分解所得的因数越来越小，但不能为1.这时每一个因数都是素因数.于是，我们证明了

定理 1　每个大于 1 的整数要么是素数，要么是若干素数的乘积.

数的这种逐步分解过程可以由多种方法实现.

例
$$60 = 4 \times 15 = 2 \times 2 \times 3 \times 5;$$
$$60 = 30 \times 2 = 15 \times 2 \times 2 = 3 \times 5 \times 2 \times 2.$$

这说明分解的方法可以是多种多样的.特别对大合数更是如此.但是一个重要的事实是，不管这种素因数分解是如何实现的，除了这些因数的次序外，所得的结果总是一样的，即在同一个数的任意两个素因数分解式中，素数是相同的，且每个素数均出现相同的次数.例如在 60 的分解中，两种分解方式得到的结果是一样的，都含有两个 2，一个 3 和一个 5.对任一数 a，这一结果可以表述为：

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n},$$

其中 $p_1, p_2, \cdots, p_n$ 是素数，$\alpha_i \geqslant 1(i = 1,2,\cdots,n)$ 是自然数.

定理 2(算术基本定理)　一个数的素因数分解式是唯一的.

注　这个定理读者是熟悉的，并且天天在用，但可能不少人并不知道它的证明，或者认为唯一分解定理是天经地义的，不必证明.事实上，这个定理不是天然成立的，需要给予证明.学了后面的内容就会明白了.

证　反证法.假设唯一分解定理不成立，那么一定存在一些数，它们具有不止一种素因数分解式.在这些数中必有一个最小的，设为 c. c 有最小素因数 p，于是有

$$c = p \cdot d.$$

因为 $d < c$，所以 d 有唯一的素因数分解式.这说明，c 的含 p 的素因数分解式是唯一的.

但是，依假设，c 至少有两种素因数分解式，所以 c 必有一种不含 p 的分解式.设在这个分解式中的最小素数是 p_1，并有

$$c = p_1 \cdot d_1. \tag{1}$$

因为 $p_1 > p$，故有 $d_1 < d$. 从而也有 $pd_1 < pd = c$. 我们来讨论数

$$c_1 = c - pd_1 = p_1 d_1 - pd_1 = (p_1 - p)d_1.$$

这是一个比 c 小的数，必有唯一分解式. 由上式，

$$p \mid c \Rightarrow p \mid c_1 \Rightarrow p \mid (p_1 - p)d_1 \Rightarrow p \mid d_1 \text{ 或 } p \mid (p_1 - p).$$

但因 p_1 是分解式(1)中的最小素数，所以 d_1 的素因数均大于 p. 这样，唯一的可能是 $p \mid (p_1 - p)$，因此它也整除 p_1. 这与 p_1 是素数矛盾. 这个矛盾说明前面的假设是不对的，即一个数不能有两种不同的分解式. 定理证毕.

在普通整数中算术基本定理成立具有重要的意义，许多定理的证明都要用到它，如$\sqrt{2}$ 是无理数的证明就用到它. 但是，算术基本定理只在普通整数中成立，在更广泛的数类中，它可以不成立.

这里顺便指出，为什么不把 1 作为素数；因为把 1 看作素数，唯一性定理就不成立了.

4.1.6 另一种"算术"

一个数只能以一种方式分解为素数的乘积，这一事实并不显然. 实际上有许多"算术"在其中类似的定理不成立. 这里给出一个简单例子. 考察全体偶数

$$2, 4, 6, 8, 10, \cdots,$$

这些数中有的可以分解为两个偶因数的乘积，而有的则不能. 不能分解为两个偶因数的乘积的偶数称为偶素数. 它们就是能被 2 整除但不能被 4 整除的偶数：

$$2, 6, 10, 14, \cdots.$$

不难看出，每一个偶数，要么是一个偶素数，要么可以表示为偶素数的乘积. 但是这样的偶素数分解式不是唯一的. 例如数 420 就有三种不同的分解式：

$$420 = 6 \cdot 70 = 10 \cdot 42 = 14 \cdot 30.$$

这种"算术"指出，素因子分解式的唯一性并不是总成立的. 在代数数论中，素因子分解式的唯一性问题涉及到费马大定理的证明.

4.1.7　最大公因数

设 a,b 是两个整数. 当我们知道了 a,b 的素因子分解式时，就不难求出它们所有的公因数. 设

$$a = p_1^{\alpha_1}\cdots p_r^{\alpha_r}, b = p_1^{\beta_1}\cdots p_r^{\beta_r}. \tag{2}$$

这里我们把两个分解式写得好像 a 和 b 有相同的素因数：

$$p_1, p_2, \cdots p_r,$$

但约定可以出现 0 指数. 例如若 p_1 整除 a 而不整除 b，则在(2)中 $\beta_1 = 0$. 例如，$a = 140, b = 110$ 就可写为

$$140 = 2^2 5^1 7^1 11^0,\ 110 = 2^1 5^1 7^0 11^1.$$

从(2)可看出，a 与 b 的公因数只是那些同时出现在 a 与 b 中的那些素因数 p_i，并且 p_i 的指数不能超过 α_i 与 β_i 中较小的一个，以及它们的乘积. 在 a,b 的所有公因数中有一个最大的，称为最大公因数. a 与 b 的最大公因数记为

$$d = (a,b).$$

令 $\gamma_i = \min(\alpha_i, \beta_i)$，则

$$d = p_1^{\gamma_1}\cdots p_r^{\gamma_r}. \tag{3}$$

例　$(140,110) = 2 \cdot 5 = 10$.

若 $d = (a,b) = 1$，则称 a 和 b 是互素的. 一个经常用到的性质是

定理 3(除法规则)　若 $c \mid ab, (c,b) = 1$，则 $c \mid a$.

证　因 $(c,b) = 1$，所以 c 的所有素因数都能整除 a，但不整除 b，并且出现在 a 的方幂不会小于出现在 c 的方幂.

4.1.8　函数$[x]$，$\{x\}$

在上一节里面我们讨论了把任意一个正整数分解成标准分解式的问题. 现在我们介绍两个数论里面常常用到的函数：$[x]$ 与 $\{x\}$.

定义　函数$[x]$ 与 $\{x\}$ 是对于一切实数都有定义的函数，函数

$[x]$ 的值等于不大于 x 的最大整数；函数 $\{x\}$ 的值是 $x-[x]$. 我们把 $[x]$ 叫做 x 的整数部分，$\{x\}$ 叫做 x 的小数部分.

例　$[\pi]=3,[e]=2,[-\pi]=-4,\left[\frac{2}{3}\right]=0,\left[-\frac{3}{5}\right]=-1$;

$$\{\pi\}=0.14159\cdots,\left\{-\frac{3}{5}\right\}=\frac{2}{5},\{\sqrt{2}\}=0.414\cdots,$$

$$\{-\pi\}=-\pi-[-\pi]=4-3.14159\cdots=0.85840\cdots.$$

$[x]$ 与 $\{x\}$ 是数学中两个十分有用的符号，下面还会不断用到. 由定义可以立刻得出下列简单性质：

1) $x=[x]+\{x\}$.

2) 若 $x\leqslant y$，则 $[x]\leqslant[y]$.

3) 若 $x=m+v$，m 是整数，$0\leqslant v<1$，则 $m=[x]$，$v=\{x\}$. 特别地，当 $0\leqslant x<1$ 时，$[x]=0$，$\{x\}=x$.

4) 对任意的整数 m 有，$[x+m]=[x]+m$，$\{x+m\}=\{x\}$.

这说明，$\{x\}$ 是周期为 1 的周期函数，它们的图形见图 4－1，图 4－2.

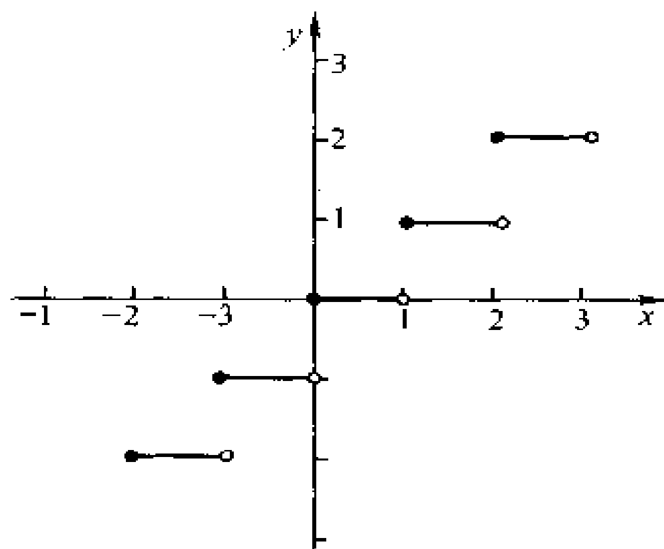


图 4－1

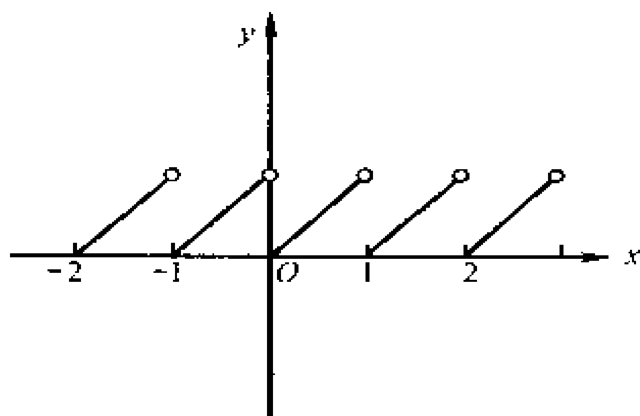


图 4 - 2

5) $[x]+[y]\leqslant[x+y]\leqslant[x]+[y]+1$，其中等号有且只有一个成立.

证　由 $x+y=[x]+[y]+\{x\}+\{y\}$，及 $0\leqslant\{x\}+\{y\}<2$.

当 $0\leqslant\{x\}+\{y\}<1$ 时，$[x+y]=[x]+[y]$.

当 $1\leqslant\{x\}+\{y\}<2$ 时，$x+y=[x]+[y]+1+(\{x\}+\{y\}-1)$，所以，$[x+y]=[x]+[y]+1$.

例　$[3.5]+[4.1]=3+4=7$；$[3.5+4.1]=[7.6]=7$；

$[3.5]+[4.6]=3+4=7$；$[3.5+4.6]=[8.1]=8$.

6) $[-x]=\begin{cases}-[x]-1, & \text{当 } x \text{ 不是整数时},\\ -[x], & \text{当 } x \text{ 是整数时}.\end{cases}$

证　当 x 是整数时，上式显然成立. 今设 x 不是整数. 我们有，

$-x=-[x]-\{x\}=-[x]-1+1-\{x\}$；$0<1-\{x\}<1$.

从而
$$[-x]=-[x]-1.$$

7) 若 a,b 是两个整数，$b>0$，则

$$a = b\left[\frac{a}{b}\right] + b\left\{\frac{a}{b}\right\}, 0 \leqslant b\left\{\frac{a}{b}\right\} \leqslant b - 1.$$

证 首先，$\frac{a}{b} = \left[\frac{a}{b}\right] + \left\{\frac{a}{b}\right\} \Rightarrow a = b\left[\frac{a}{b}\right] + b\left\{\frac{a}{b}\right\}$.

其次，a 是整数，$b\left[\frac{a}{b}\right]$是整数，从而两者之差 $b\left\{\frac{a}{b}\right\}$ 是整数，且小于 b. 所以

$$0 \leqslant b\left\{\frac{a}{b}\right\} \leqslant b - 1.$$

8) 若 a, b 是任意两个正整数，则不大于 a 而为 b 的倍数的正整数的个数是$\left[\frac{a}{b}\right]$个.

例 取 $a = 10, b = 3$，则$\left[\frac{10}{3}\right] = \left[3 + \frac{1}{3}\right] = 3$. 在$[0, 10]$含 3 个 3 的倍数.

证 当 $a < b$ 时显然. 设 m 是任一不大于 a 而为 b 的倍数的正整数，则

$$0 < m = bm_1 \leqslant a, 0 < m_1 \leqslant \frac{a}{b}.$$

故满足以上条件的 m 的个数等于满足上式的 m_1 的个数，因而等于$\left[\frac{a}{b}\right]$. 证完.

4.1.9 费马素数

一种有趣且有很长历史的数叫费马素数. 这些数是由法国数学家费马引进的. 最初的五个费马素数是

$F_0 = 2^{2^0} + 1 = 3$, $F_1 = 2^{2^1} + 1 = 5$, $F_2 = 2^{2^2} + 1 = 17$,

$F_3 = 2^{2^3} + 1 = 257$, $F_4 = 2^{2^4} + 1 = 65537$.

由这些数可以看出，费马素数的一般公式是

$$F_n = 2^{2^n} + 1. \tag{4}$$

尽管除了上面的五个数外，费马没有做进一步的计算，但他坚信所有的这种数都是素数. 然而当瑞士数学家欧拉再往前走了一步，这个猜想就被推翻了. 他证明了下一个费马数不是素数：

$$F_5 = 4294967297 = 641 \cdot 6700417.$$

到 1988 年时，数学家已经知道，$F_6, F_7, \cdots, F_{21}$ 都是合数.

故事到此并没有结束，费马数又出现在用直尺和圆规作正多边形的这样一个完全不同的问题中.

正多边形是这样一种多边形，它的顶点等距离地位于一个圆周上. 如果它有 n 个顶点，就称它为正 n 边形. 从顶点到圆心的 n 条连线构成 n 个中心角，每个角为 $360°/n$. 如果能够作出这样大小的一个角，就能作出这个正 n 边形.

古希腊人对于寻找用直尺和圆规作正多边形的方法十分感兴趣. 自不待言，对于等边三角形与正方形这种简单情形他们是会作的. 利用不断平分中心角的办法，他们能够作出

$$4, 8, 16, 32, \cdots, 2^n, \cdots$$

$$3, 6, 12, 24, \cdots, 3 \cdot 2^n, \cdots$$

个顶点的正多边形. 此外他们还能作正五边形. 因此，也能作出

$$5, 10, 20, 40, \cdots, 5 \cdot 2n, \cdots$$

个顶点的正多边形. 这样一来，他们又可作出另一种正多边形. 正十五边形的中心角是

$$360°/15 = 24°.$$

而这可由正五边形的中心角 72° 及正三边形的中心角 120° 来作出：第一个角的两倍减去第二个角：$72° \times 2 - 120° = 24°$. 因此他们又能作出边数为

$$15, 30, 60, \cdots, 15 \cdot 2^n, \cdots$$

的正多边形.

以上就是希腊人在构造正 n 边形上所作出的贡献. 他们的贡献保持了两千年的时间，直到年青的德国数学家高斯 1801 年发表了数

论的划时代著作《算术研究》,这个问题才有了新的进展.高斯超过希腊人的不仅仅是他给出了一个利用直尺和圆规来作正十七边形的方法,更重要的是,对所有的 n 他解决了哪些正 n 边形可以用直尺和圆规作出来,而哪些则不能.下面我们来叙述高斯的结果.

上面已经指出,从一个正 n 边形出发,通过等分它的每个中心角,就能得到正 $2n$ 边形.另一方面,从一个正 $2n$ 边形出发,只要取 n 个不相邻的顶点就能得到正 n 边形.这表明,为了判定哪些正 n 边形可作,只要讨论奇数情形就够了.高斯证明了

定理 4　对奇数 n,当且仅当 n 是一个费马素数,或是若干个不同的费马素数的乘积时,正 n 边形才能用直尺和圆规作出来.

让我们考察几个最小的值 n.正 3 边形和正 5 边形可以作出,但不能作出正 7 边形,因为 7 不是费马素数.也不能作出正 9 边形,因为 $9 = 3 \cdot 3$ 是两个相等的费马素数的乘积.也不能作出 $n = 11$ 和 $n = 13$ 的正多边形,但是能够作出能 $n = 15 = 3 \cdot 5$ 及 $n = 17$ 的正 n 边形.

很自然,高斯的发现引起人们对费马素数的新兴趣.迄今没有新的费马素数被发现.数学家倾向于相信不再有其它的费马素数了.

4.1.10　完全数与梅森数

在数论的发展史上充满了著名猜想和未解难题.这一节我们将注意力集中于与完全数和梅森数有关的猜想上.其中少数猜想得到了满意的解答,但大部分猜想还没有解决.梅森(Mersenne Marin,1588 – 1648)是法国数学家,自然哲学家和宗教家.他在 1644 年提出了梅森素数.梅森素数的提出是探索表素数公式的开始,在数论史上具有开拓性的意义.

定义　形如

$$M_n = 2^n - 1 \qquad (n > 1)$$

的数叫做梅森数,其中是素数的梅森数叫做梅森素数.

例 $M_2 = 2^2 - 1 = 3, M_3 = 2^3 - 1 = 7, M_4 = 2^4 - 1 = 15, M_5 = 2^5 - 1 = 31$.

梅森提出的问题具有启发性，但他当时的判断有误. 他说，对 $p = 2,3,5,7,13,17,31,67,127,257$, M_p 是素数. 而 $p < 257$ 的其它素数对应的 M_p 都是合数. 梅森是如何得到这一结论的呢？无人知晓. 到了1947年有了台式计算机后，人们才能检查他的结论. 发现他犯了五个错误. M_{67}, M_{257} 不是素数，而 M_{61}, M_{89}, M_{107} 是素数.

1867年以来，人们已经知道 M_{67} 是合数，但对它的因子一无所知. 1903年10月在美国数学会举行的一次会上，数学家科尔(Fredrick Nelson Cole)提交一篇论文《大数的因子分解》. 轮到科尔报告时，他走到黑板前，一言未发便作起2的方幂的演算，直到2的67次幂，从所得结果减去1. 然后默默无言地在黑板的空白处写下两个数相乘：

$$193707721 \times 761838257287.$$

两个计算结果完全一样. 之后，他只字未吐又回到自己的座位上. 会场上爆发了热烈的掌声.

关于梅森数我们有下面的定理.

定理 5 若 $n > 1$，且 $a^n - 1$ 是素数，则 $a = 2$，且 n 是素数.

证 先证 a 必须是2. 设 $a > 2$. 因为 $n > 1$，所以

$$1 < a - 1 < a^n - 1 = (a - 1)(a^{n-1} + \cdots + a + 1)$$

所以 $a^n - 1$ 有大于1的真因数 $(a - 1)$. 这就是说 $a^n - 1$ 不是素数，因此要它是素数必有 $a = 2$. 这就是说，$a^n - 1$ 是素数，它必须是 $2^n - 1 = M_n$ 的形式，即必须是梅森数.

再证 n 一定是素数. 事实上，如果 n 是合数：$n = kl$，其中 $1 < k < n$，则必有 $1 < 2^k - 1 < 2^n - 1$. 于是，$(2^k - 1) \mid (2^n - 1) = (2^{kl} - 1)$，从而 $2^n - 1$ 不是素数. 这样一来，要 $2^n - 1$ 是素数，必须 n 是素数. 这说明 $a^n - 1$ 是素数，则它一定是梅森素数. 证毕.

这个定理只是给出了 $a^n - 1$ 是素数的必要条件. 这个条件不是

充分的. 换言之, M_n 是素数必须 n 是素数, 但反过来并不成立: 当 n 是素数时, M_n 不一定是素数. 前面指出, M_{67}, M_{257} 都不是素数. 其它例子有:

$$23 \mid M_{11}, 47 \mid M_{23}, 167 \mid M_{83}, 263 \mid M_{131}, 359 \mid M_{179}$$

等等.

那么到底有多少梅森数是素数呢?

到今天为止, 我们只知道 34 个梅森数是素数. 它们是 M_p, 其中

$$p = 2,3,5,7,13,17,19,31,61,89,107,127,521,607,$$
$$1279,2203,2281,3217,4253,4423,9689,9941,$$
$$11213,19937,21701,23209,44497,86243,110503,$$
$$132049,216091,756839,858433,1257787.$$

从第 13 个开始, 即从 M_{521} 开始, 都是在 1952 年以后, 借助电子计算机而陆续发现的. 目前所知道的最大素数, 就是梅森素数 $M_{1257787}$, 这个数有 378632 位. 这是 1996 年 5 月美国威斯康星州克雷研究所发现的.

梅森数中是否有无穷多个素数? 这是一个没有解决的问题.

还有人提出过这样的猜想, 即如果 M_p 是素数, 那么 M_{M_p} 也是一个素数. 这个猜想对于小的梅森数都是对的. 但到第 5 个梅素数 $M_{13} = 8191$, 这个猜想就被否定了. 借助于电子计算机, 可以证明 $M_{M_{13}} = 2^{8191} - 1$ 是一个合数. 这个数有 2466 位, 到 1976 年, 才找到它的一个素因子

$$p = 2 \times 20644299 \times M_{13} + 1$$
$$= 338193759479.$$

到 1957 年, 有人证明了虽然 M_{17} 与 M_{19} 都是素数, 但 $M_{M_{17}}$ 与 $M_{M_{19}}$ 都是合数, 它们可以分别被 $1768(2^{17} - 1) + 1$ 与 $120(2^{19} - 1) + 1$ 整除. 已知最大的梅森复合数为 M_q, 其中

$$q = 39051 \times 2^{6001} - 1.$$

与梅森数密切相关的是寻找偶完全数的问题. 前面已经说过，完全数的概念来自毕达哥拉斯. 他把完全数定义为等于自己的真因数之和的自然数. 例如6是完全数，因为 $6=1+2+3$. 下一个完全数是28，因为28的真因子是1,2,4,7,14，而 $28=1+2+4+7+14$. 现在的数论书中对完全数的概念作了一点修改，将完全数定义为

定义 一个自然数 n 称为完全数，如果它的全部因数之和等于 $2n$.

我们用 $\sigma(n)$ 表示 n 的全部因数之和. 于是我们有

$$\sigma(1)=1, \sigma(2)=3, \sigma(3)=4, \sigma(4)=7, \sigma(5)=6,$$

$$\sigma(6)=1+2+3+6=12=2\cdot 6,$$

$$\sigma(28)=1+2+4+7+14+28=56=2\cdot 28.$$

为了下面使用方便，我们先给出 $\sigma(n)$ 的计算公式：

设 $n=p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$. 为了计算 $\sigma(n)$ 考虑下面的乘积，

$$(1+p_1+p_1^2+\cdots p_1^{k_1})(1+p_2+p_2^2+\cdots+p_2^{k_2})\cdots(1+p_r+p_r^2+\cdots+p_r^{k_r}).$$

在这个乘积的展式中，n 的每一个因子都出现一次，而且只出现一次，所以

$$\sigma(n)=(1+p_1+p_1^2+\cdots p_1^{k_1})(1+p_2+p_2^2+\cdots+p_2^{k_2})\cdots(1+p_r+p_r^2+\cdots+p_r^{k_r}).$$

利用几何级数求和公式，得

$$\sigma(n)=\frac{p_1^{k_1+1}-1}{p_1-1}\frac{p_2^{k_2+1}-1}{p_2-1}\cdots\frac{p_r^{k_r+1}-1}{p_r-1}.$$

有了这一公式我们就可以证明下面的引理.

引理 若 $(m,n)=1$，则 $\sigma(mn)=\sigma(m)\sigma(n)$.

证 若 $m=p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}, n=q_1^{j_1}q_2^{j_2}\cdots q_s^{j_s}$ 则由计算公式

$$\sigma(mn)=\left[\frac{p_1^{k_1+1}-1}{p_1-1}\cdots\frac{p_r^{k_r+1}-1}{p_r-1}\right]\left[\frac{q_1^{j_1+1}-1}{q_1-1}\cdots\frac{q_s^{j_s+1}-1}{q_s-1}\right]$$

$$= \sigma(m)\sigma(n).$$

现在再回到完全数的问题.古希腊人已经知道 4 个完全数.活动于公元一世纪的新毕达哥拉斯学派的哲学家和数学家尼科玛科斯(Nicomachus)在他的《算术引论》中列出了下面 4 个完全数:

$$P_1 = 6,\ P_2 = 28,\ P_3 = 496,\ P_4 = 8128.$$

他说,这些数的形成很有规律:个位数有一个,十位数有一个,百位数有一个,千位数有一个.只根据这几个有限的数据,他就作了如下两个猜想:

1)第 n 个完全数恰有 n 位数字;

2)偶完全数总是交替地以 6 和 8 结尾.

可惜这两个猜想都是错误的.不存在 5 位数字的完全数.下一个完全数是 15 世纪发现的:

$$P_5 = 33550336.$$

它的末位数字是 6.第 6 个完全数是

$$P_6 = 8589869056$$

末位数字也是 6,而不是 8.借助同余的理论可以证明,偶完全数总是以 6 和 8 结尾,但不一定交替出现.P_6 已经这样大,似乎告诉我们完全数十分稀少.现在我们还不知道有无限个完全数,还是只有有限个完全数.

关于确定所有完全数的一般形式的问题要追溯到数学的发祥时期.欧几里得已经部分地解决了这一问题.他证明了如果和

$$1 + 2 + 2^2 + \cdots 2^{k-1} = p$$

是素数,则 $2^{k-1}p$ 是一个完全数.例如,$1 + 2 + 4 = 7$ 是一个素数,因此 $4 \times 7 = 28$ 是一个完全数.欧几里得的推理用了几何级数的求和公式:

$$1 + 2 + 2^2 + \cdots 2^{k-1} = 2^k - 1.$$

换句话说,如果 $2^k - 1(k > 1)$ 是素数,则 $n = 2^{k-1}(2^k - 1)$ 是一个完全数.大约在欧几里得 2000 年后,欧拉证明了,所有的偶完全数一

定具有这种形式.我们把这两个结论并在一起,得出下面的定理.

定理 6 如果 M_p 是素数,那么

$$\frac{1}{2}M_p(M_p+1)=2^{p-1}(2^p-1) \tag{5}$$

是一个偶完全数,而且除这些以外,再没有其它的偶完全数.

证 证明分为两步:1)(5)是完全数;2)偶完全数都具有(5)的形式.

1)根据完全数的定义,只需证明:如果 $M_p=2^p-1$ 是素数,那么

$$\sigma(2^{p-1}(2^p-1))=2\cdot 2^{p-1}(2^p-1)=2^p(2^p-1).$$

注意到$(2^{p-1},2^p-1)=1$,所以

$$\sigma(2^{p-1}(2^p-1))=\sigma(2^{p-1})\sigma(2^p-1).$$

2^{p-1} 的因数显然为 $1,2,2^2,\cdots,2^{p-1}$,所以

$$\sigma(2^{p-1})=1+2+\cdots+2^{p-1}=2^p-1.$$

这个计算还说明,2^{p-1} 不是完全数.

因为 $M_p=2^p-1$ 是素数,它只有两个因数 2^p-1 和 1,所以

$$\sigma(2^p-1)=2^p-1+1=2^p$$

这样一来,$\sigma(2^{p-1}(2^p-1))=2^p(2^p-1)$.这就证明了(5)是完全数.

2)任一偶完全数都具有(5)的形式.现在假定 a 是一个偶完全数,并假设 a 的标准分解式中含2的最高方幂的次数为 $n-1$.因 a 为偶数,所以 $n-1\geqslant 1$.又因 2^{n-1} 不是偶完全数,所以 $a=2^{n-1}u$,$u>1$,$2\nmid u$.因此 a 的因数为所有形如 2^iv 的数,其中 $0\leqslant i\leqslant n-1$ 及 $v\mid u$.从而

$$\begin{aligned}2^nu=2a=\sigma(a)&=\sigma(2^{n-1}u)=\sigma(2^{n-1})\sigma(u)\\&=(1+2+\cdots 2^{n-1})\sigma(u)\\&=\sigma(u)(2^n-1).\end{aligned}$$

即得$(2^n-1)\mid 2^nu$.因$(2^n,2^n-1)=1$,所以$(2^n-1)\mid u$.即$\frac{u}{2^n-1}$是

整数. 另一方面,由上面的等式得到

$$\sigma(u)=\frac{2^n u}{2^n-1}=u+\frac{u}{2^n-1}.$$

但 u 与 $\frac{u}{2^n-1}$ 都是 u 的因数,而 $\sigma(u)$ 又是 u 的所有因数的总和,所以 u 只有两个因数 u 和 $\frac{u}{2^n-1}$. 因 $u>1$ 及 u 至少有两个因数 u 与 1,所以必须 $\frac{u}{2^n-1}=1$. 换句话说,u 是一个素数,且

$$u=2^n-1$$

由定理 5,n 必须是素数. 这就证明了 $a=2^{n-1}u=2^{n-1}(2^n-1)=\frac{1}{2}M_n(M_n+1)$,$n$ 是素数,证毕.

当 $p=2,3,5,7$ 时,2^p-1 的值分别是 3,7,31,127,它们都是素数,所以

$$2(2^2-1)=6,\qquad 2(2^3-1)=28,$$
$$2^4(2^5-1)=496,\qquad 2^6(2^7-1)=8128.$$

都是完全数.

这个定理说明,是否有无穷多个偶完全数的问题归结为是否有无穷多个梅森素数的问题. 由于目前只知道 34 个梅森素数,所以目前只知道 34 个偶完全数,其中最大的是 $2^{1257786}(2^{1257786}-1)$. 是否存在奇完全数?这是一个没有解决的问题. 借助计算机可以证明:1) 若 n 为奇完全数,则 $n>10^{300}$;2) 若 n 为奇完全数,则 n 必有一个大于 100110 的素因数;3) 若 n 为奇完全数,则 n 的互异的素因数的个数至少是 8.

4.1.11　高斯的功绩

讲到数论不能不提到德国数学家高斯(Gauss Carl Friedrich, 1777 – 1855). 他与阿基米德和牛顿并列为历史上最伟大的数学家. 他不仅对纯粹数学作出意义深远的贡献,而且对天文学、大地测量和

电磁学也作出重要贡献.

1795－1798年高斯在格丁根大学学习,1799年获该大学博士学位.博士论文的题目是代数基本定理的一个证明,而在他之前的证明都是不完全的.24岁时他发表了《算术研究》,这是数学史上最出色的成果之一.书中广泛而系统地阐述了数论里重要的概念和方法.他研究了同余的理论,给出了二次互反定律的第一个证明.他利用数论对正 n 边形作图问题提出了代数解法.正如前面刚讲过的,对只用直尺圆规作正 n 边形的作图问题,他给出了判别准则.特别是,高斯证明了能够作出正17边形,他自己就完成了这个作图.这个发现是欧几里得后的第一个.他很以此为骄傲.因此,遵照他的遗嘱,在他的墓碑上画了一个内接于圆的正17边形.他使用了复数,并在1831年借助复数的平面表示建立了严密的复数理论.高斯是首先认识到非欧几何存在的人.高斯奠定了曲面内蕴几何学的基础,启发他的学生B.黎曼发展了高维空间的一般的内蕴几何,这在后来成了爱因斯坦的广义相对论的数学基础.

高斯既是对数字具有非凡记忆力的卓越的算术家,又是一位渊博的理论家和杰出的应用数学家.他通过大约有155个题目的出版物促进了数学的发展.他立论极端谨慎,有3个原则:"少些,但是要成熟";"不留下进一步要做的事情";极度严格的要求.从他逝世后出版的著作可以明显地看到,他有大量重要的文章从未发表.由于他在数学,天文学,大地测量学和物理学中的杰出的研究工作,他被选为许多科学院和学术团体的成员."数学家之王"的称号是对他极其恰当的赞誉.

习 题

1. 造不超过100的素数表.
2. 若 $n^2 = ab$,$(a, b) = 1$,则 a, b 都是平方数.

3. 对任意整数 a，证明

$2\mid a(a+1),\quad 3\mid a(a+1)(a+2),\quad 3\mid a(2a^2+7).$

4. a 是奇数则 $32\mid(a^2+3)(a^2+7)$.

5. 对任意整数 a，$(2a+2,9a+4)=1$.　（用辗转相除法.）

6. 若 a 是奇数，则 $(3a,3a+2)=1$.

7. 任何素数的幂都不是完全数.

8. 一个平方数不会是完全数.

9. 两个奇素数的乘积不会是完全数.

10. 证明数 $n=2^{10}(2^{11}-1)$ 不是一个完全数.

§4.2　素数定理与哥德巴赫猜想

4.2.1　素数定理

关于素数的第一个问题是，有多少个素数？这个问题约在 2300 年前已由希腊人回答了. 其解答见于欧几理得的《几何原本》第九卷定理 20.

定理 1　素数有无穷多个.

证　欧几里得的证法是反证法. 假定素数只有有限个，将它们罗列如下：

$$p_1=2,p_2=3,\cdots,p_N.$$

那么数 $p_1p_2\cdots p_N+1$ 将不为上述素数中任一个所整除. 因此，或者它本身是素数，或者它有不同于上述素数的新的素因子. 这与假定矛盾. 定理证毕.

欧几里得关于素数是无限的证明是数学证明中的一个典范. 如果世界上确实有经典性的伟大定理的话，那么欧几里得的证明就是一例. 实际上，他的论证常常被人们作为数学定理的典范. 因为这一定理简洁，优美，又极为深刻. 20 世纪英国数学家哈代在其精彩的专

著《一个数学家的辩白》中对欧几里得的证明作了如下的评论:"我最好还是回到古希腊人那里去.我要叙述并证明希腊数学中两个有名的定理.这两个定理都很简单","在思想和演算上都很简单,但毫无疑问它们是最高水平的定理.每一个定理现在仍然像它们刚发现时那样生气勃勃而举足轻重——两千年的岁月没有使它们产生一丝陈旧感".哈代说的另一个定理是关于$\sqrt{2}$是无理数的证明.

由欧几里得的证明我们引出另一个有趣的问题.对一个素数p,用$p^{\#}$表示所有小于等于p的素数的乘积.我们把形如$p^{\#}+1$的数叫做欧几里得数,因为这些数出现在素数是无穷的欧几里得证法中.有趣的是,前五个欧几里得数都是素数:

$$2^{\#}+1=2+1=3,$$

$$3^{\#}+1=2\times3+1=7,$$

$$5^{\#}+1=2\times3\times5+1=31,$$

$$7^{\#}+1=2\times3\times5\times7+1=211,$$

$$11^{\#}+1=2\times3\times5\times7\times11+1=2311.$$

但是

$$13^{\#}+1=59\times509,$$

$$17^{\#}+1=19\times97\times277,$$

$$19^{\#}+1=347\times27953$$

都不是素数.是否存在无穷多个素数p,使$p^{\#}+1$是素数,这个问题还没有解决.是否存在无穷多个素数p,使$p^{\#}+1$是合数,这个问题也没有解决.

关于素数的第二个问题是,相邻素数的间距有多大?

定理 2 相邻素数的间距要多大有多大.

证 通过举例证明.存在 999 个连续的自然数,其中没有一个是素数.它们是

$$1000!+2,\ 1000!+3,\ 1000!+4,\ \cdots,1000!+1000.$$

易见,第一个数能被 2 整除,第二个数能被 3 整除,…,最后一个

数能被 1000 整除．这就造出了 999 个连续的自然数，其中没有一个是素数．

同样的办法可以造出更大的间隔．这就完成了定理的证明．

今问有多少对相差为 1 的素数？很清楚，2 是唯一的偶素数．其它素数都是奇素数，它们的差是偶数．这样一来，2 与 3 是唯一的一对相差为 1 的素数．同样地，2 和 5 是唯一的一对相差为 3 的素数，2 和 7 是唯一的一对相差为 5 的素数．不存在相差为 7 的一对素数．

相差为 2 的素数怎样呢？显然，这样的素数必需都是奇数．这种素数对叫做孪生素数．如

$$3,5;\ 5,7;\ 11,13;\ 17,19;\ 29,31;\cdots.$$

孪生素数对的个数是有限，还是无限？这个问题仍然没有答案．半个世纪以来，这已成为数论中最高深的研究课题之一了．

下面研究素数分布的情况．

考察一下素数表就会发现，素数的出现是无规则的．但是，素数的无序后面隐藏着有序．最终，素数的间隔也表现出某种秩序．

1791 年高斯通过对素数表的调查注意到，虽然间隔呈无序状态，但平均间隔增长得很慢．他猜测到如下的素数定理，但没有证明．

用 $\pi(x)$ 表示不超过 x 的素数的个数．观察下表：

x	$\pi(x)$	$\frac{x}{\ln x}$	$\frac{\pi(x)\ln x}{x}$
1000	168	145	1.16
10000	1229	1086	1.13
50000	5133	4621	1.11
100000	9592	8686	1.10
500000	41538	38103	1.090
1000000	78498	72382	1.084
2000000	148933	137848	1.080
5000000	348513	324149	1.075
10000000	664579	620417	1.071

这个表告诉我们如下几点：

1）素数的个数是无穷的：$\pi(x) \to \infty$；

2）素数的个数与全体整数的个数相比则少得多：$\pi(x)/x \to 0$；即，几乎所有的整数都不是素数.

3）$x/\log x$ 是 $\pi(x)$ 的渐近式.

因而，人们猜想下面的定理成立：

定理 3（素数定理） $\lim \dfrac{\pi(x)}{\dfrac{x}{\log x}} = 1.$

首先对这个问题作出重要贡献的是切比雪夫.他在 1848 年和 1850 年证明了：

定理 4 $\left(\dfrac{\log 2}{3}\right)\dfrac{x}{\log x} < \pi(x) < (6\log 2)\dfrac{x}{\log x}.$

1896 年，法国数学家阿达马和泊松几乎同时相互独立地证明了素数定理.但他们使用了复变函数的相当深入的理论.这就推动人们去寻找初等的和较简单的证明.直到 1949 年，塞尔伯格和厄尔都斯才给出了既不用复变函数也不用微积分的证明.这曾在数学界引起惊奇，因为数学家花了一个半世纪才找到它.

4.2.2 哥德巴赫猜想

1742 年，德国数学家哥德巴赫（Christian Goldbach，1690－1764）在和他的好朋友、大数学家欧拉（Leonhard Euler，1707－1783）的几次通信中，提出了关于正整数和素数之间关系的两个推测，用现在确切的话来说，就是：

（A）每一个不小于 6 的偶数都是两个奇素数之和；

（B）每一个不小于 9 的奇数都是三个奇素数之和.

这就是著名的哥德巴赫猜想.我们把猜想（A）称为"关于偶数的哥德巴赫猜想"，把猜想（B）称为"关于奇数的哥德巴赫猜想".由于

$$2n + 1 = 2(n - 1) + 3$$

所以，从猜想（A）的正确性立即推出猜想（B）的正确性.

为了增加一些感性认识,我们给出 30 以内的偶数的表示:

$6 = 3 + 3$,　　$8 = 3 + 5$,

$10 = 3 + 7 = 5 + 5$,　　$12 = 5 + 7$,

$14 = 3 + 11 = 7 + 7$,　　$16 = 3 + 13 = 5 + 11$,

$18 = 5 + 13 = 7 + 11$,　　$20 = 3 + 17 = 7 + 13$,

$22 = 3 + 19 = 5 + 17 = 11 + 11$,

$24 = 5 + 19 = 7 + 17 = 11 + 13$,

$26 = 3 + 23 = 7 + 19 = 13 + 13$,　　$28 = 5 + 23 = 11 + 17$,

$30 = 7 + 23 = 11 + 19 = 13 + 17$.

欧拉虽然没有能够证明这两个猜想,但是对它们的正确性是深信不疑的.1742 年 6 月 30 日,在给哥德巴赫的一封信中他写道:我认为这是一个肯定的定理,尽管我还不能证明出来.哥德巴赫猜想提出到今天已经有 258 年了,可是至今还不能最后地肯定它们的真伪.人们积累了许多宝贵的数值资料,都表明这两个猜想是合理的.这种合理性以及猜想本身所具有的极其简单、明确的形式,使人们和欧拉一样,也不由得不相信它们是正确的.因而,二百多年来这两个猜想一直吸引了许许多多数学工作者和数学爱好者,特别是不少著名数学家的注意和兴趣,并为此作出了艰巨的努力.

从提出哥德巴赫猜想到 19 世纪结束这 160 年中,虽然许多数学家对它进行了研究,但并没有得到任何实质性的结果和提出有效的研究方法.这些研究大多是对猜想进行数值的验证,提出一些简单的关系式或一些新的推测.1900 年,在巴黎召开的第二届国际数学会上,德国数学家希尔伯特(D. Hilbert)在其展望 20 世纪数学发展前景的著名演讲中,提出了 23 个他认为是最重要的没有解决的数学问题,作为今后数学研究的主要方向,并期待在这新的一个世纪里,数

学家们能够解决这些难题.哥德巴赫猜想就是希尔伯特所提出的第八问题的一部分.但是,在此以后的一段时间里,对哥德巴赫猜想的研究并未取得什么进展.1912年,德国数学家朗道(E. Landau)在英国剑桥召开的第五届国际数学会上十分悲观地说:即使要证明下面较弱的命题(C),也是当代数学家所力不能及的:

(C) 存在一个正整数 k,使每一个大于等于2的整数都是不超过 k 个素数之和.

1921年,英国数学家哈代(G. H. Hardy)在哥本哈根数学会作的一次讲演中认为:哥德巴赫猜想可能是没有解决的数学问题中的最困难的一个.

就在一些著名数学家作出悲观预言和感到无能为力的时候,他们没有料到,或者没有意识到对哥德巴赫猜想的研究正在开始从几个不同方向取得了为以后证明是重大的突破.

1937年,苏联数学家维诺克拉多夫(Vinogradov. A. I)证明了:每一个充分大的奇数都是三个奇素数之和.巴雷德金算过,当奇数 $n > \exp\{e^{16.038}\}$ 时,就能表为三个奇素数之和.这就是说,除掉有限个奇数外,命题(B)都成立.但 $\exp\{e^{16.038}\}$ 这个数太大了,无法逐一验证对于小于它的奇数来说命题(B)是否都成立.所以命题(B)基本上被证明了.

1938年,我国著名数学家华罗庚及一些国外数学家独立地证明了,命题(A)对几乎所有的偶数都成立.假设 $M(x)$ 表示不超过 x,而又不能表示为两个素数之和的偶数的个数,那么

$$\lim \frac{M(x)}{x} = 0.$$

这说明,使命题(A)成立的偶数出现的'概率'为1.

殆素数指素因数个数不超过某一常数的自然数.今引入下面两个命题:

(D) 每一个充分大的偶数都是素因数个数分别不超过 a 和 b 的

两个殆素数之和,记为(a,b).

(E) 每一个充分大的偶数都可以表示为一个素数与一个素因数个数不超过c的殆素数之和,记为$(1,c)$.

在命题(D)中取$a=1$,即得命题(E).但因处理这两个命题的方法有差异,所以还是分开写好.处理命题(E)需要高深的分析工具.证明哥德巴赫猜想就是要证明(1,1)成立.

布朗1920年证明了(9,9):

定理5(布朗) 每一个充分大的偶数都可以表示为素因数个数不超过9的两个殆素数之和.

关于命题(E)瑞尼1948年证明了$(1,c)$:

定理6(瑞尼) 存在一个正常数c,使每一个充分大的偶数都可以表为一个素数与一个不超过c的殆素数之和.

不少数学家改进了布朗和瑞尼的结果.我国数学家对此作出重大贡献.

1956年王元证明了(3,4).同年阿·维诺克拉多夫证明了(3,3).1957年王元又证明了(2,3).

关于命题(E),1962年潘承洞证明了(1,5).1963年潘承洞与巴尔巴恩又分别独立证明了(1,4).1965年,阿·维诺克拉多夫,布赫夕塔布与朋比尼都证明了(1,3).

1966年,我国著名数学家陈景润对筛法作了新的重要改进之后证明了(1,2):

定理7 每一个充分大的偶数都可以表为一个素数与一个不超过两个素数的乘积之和.

这是一个十分杰出的成就.

从1937年维诺克拉多夫的工作开始到现在已经60多年了.这期间哥德巴赫猜想取得了巨大的进展.尽管如此,人们还不能预测哥德巴赫猜想解决的最后日程.研究哥德巴赫猜想产生的方法不仅对数论有广泛的应用,对数学的其它分支也有广泛的应用.

4.2.3　有关素数的 12 个问题

美国数学家阿波斯托尔在 1976 年出版的《解析数论导引》一书中给出下面 12 个未解问题：

1) 是否存在大于 2 的偶数，不是两个素数之和？

2) 是否存在大于 2 的偶数，不是两个素数的差

3) 是否存在无穷多对孪生素数？

4) 是否存在无穷多个梅森素数？

5) 是否存在无穷多个梅森数是复合数？

6) 是否存在无穷多个费马素数？

7) 是否存在无穷多个费马数是复合数？

8) 是否存在无穷多个素数具有 x^2+k 的形式，其中 x 是整数？

9) 是否存在无穷多个素数具有 x^2+k 的形式，其中 k 是给定的整数？

10) 对每一个整数 n，是否在 n^2 与 $(n+1)^2$ 之间都至少存在一个素数？

11) 对每一个整数 $n>1$，是否在 n^2 与 n^2+n 之间都至少存在一个素数？

12) 是否有无穷多个素数，其每一位都是 1（如 11，11111111111111111）？

第五章　从勾股定理到费马大定理

我们感到有可能和比我们水平高许多的数学接触,这种数学的力量和美尽管只能简单地一瞥,也构成了丰富我们思想的基础,并在我们作为数学使用者和数学教师的朴素活动中给了我们长期反省的机会.

L. Felix

引　言

这一讲包含三部分内容:二元一次不定方程,勾股定理及有关问题和费马大定理.涉及到一次不定方程,二次不定方程与高次不定方程.这三个问题跨越了两千年的历史.不定方程的问题古老而常青,并且非常困难,现在仍然是数学研究的热门课题.

数论从一开始就讨论各式各样的确定方程和不确定方程.在确定方程中,未知数的个数与方程的个数一样多.例如

$$2x^2 + 3x + 1 = 0$$

和

$$\begin{cases} 2x + 3y = 1 \\ 3x + 4y = 2 \end{cases}$$

都属于确定方程.初等代数、高等代数、线性代数研究的都是确定方程.

不定方程是这样一种方程,其中变量的个数多于方程的个数,并且未知数还要受到某种限制,例如限制未知数为整数、有理数等.不定方程的讨论非常复杂.我们只讨论三个著名问题:百鸡问题、勾股定理和费马大定理.不定方程起源很早,大约在公元3世纪就有丢番

图的研究.我国《周髀算经》的商高定理"勾三股四弦五"就属于此类问题.商高定理远在丢番图之前.《周髀算经》是公元前1世纪的著作,是一部讲盖天学说的天文著作,书中有较复杂的开方和分数运算.在《周髀算经》的第一章里叙述了西周开国时期(约公元前1000年)周公姬旦与商高的问答,讨论用矩测量的方法.商高对周公说:"故折矩以为句广三,股修四,径隅五(勾的古写为句)".可见"勾三股四弦五"这个特殊例子的发现是很早的.

在西方不定方程的研究起源于丢番图.丢番图是数学史上一位杰出的数学家,他写了三部著作:1)《算术》原有13卷,现存6卷;2)《论多边形》现存一些片断;3)《衍论》已遗失.

《算术》是一部具有高度创造性的伟大著作,对欧洲的数论产生了深远的影响.书中现存部分大约有130个问题的解法.其中大部分是不定方程和不定方程组.因此,不定方程在西方也叫丢番图方程.

对丢番图方程我们主要研究些什么呢?

1)解的存在性问题.初等办法是凑一个,找一个.高级的办法是借助分析,代数或其它工具去找解.如果解不存在,则需给出证明.这里顺便指出,不存在性证明最困难.如五次以上的代数方程的求根问题、几何作图三大难题无解的证明、本章的费马大定理等都是.

如果解存在,下面的问题就是:

2)解的个数问题.

再进一步的问题是:

3)确定解的完全组.

有时还需要了解

4)解的界的估计.

能找出解的完全组这是最理想的.但这是很罕见的情况.

§5.1　一次不定方程

公元5世纪我国古代数学家张丘建在他的《算经》中就提出并解答了一个二元一次不定方程的问题.张丘建的生卒年代已不可考.他的著书的时代大约是在魏的天赐元年(公元404年)与太和十八年(公元494年)之间.

《张丘建算经》卷下最后一题是世界有名的百鸡问题:

"鸡翁一,值钱五,鸡母一,值钱三,鸡雏三,值钱一.百钱买百鸡.问鸡翁母雏各几何?"

设用 x,y,z 分别代表鸡翁、鸡母、鸡雏的数目,就得到下面的方程:

$$5x+3y+\frac{1}{3}z=100,$$

$$x+y+z=100.$$

消去 z,再化简,得到

$$7x+4y=100.$$

我们要解决这个问题,就是要求出上述方程的非负整数解.

但是上述方程不过是二元一次不定方程的一个具体的例子.二元一次不定方程的一般形式是

$$ax+by=c,$$

其中 a,b,c 是整数.下面我们研究它的解法.

5.1.1　通解公式

二元一次不定方程的任何一个具体的解都叫作它的一个特解.先假定二元一次不定方程有一个特解,我们来说明如何借这一特解将它的全部解表示出来.

定理1　设二元一次不定方程

$$ax+by=c \tag{1}$$

(其中 a,b,c 是整数且 a,b 都不是0)有一组整数解 $x=x_0,y=y_0$;又设 $(a,b)=d,a=a_1d,b=b_1d$,则(1)的一切解可以表成

$$x=x_0-b_1t,y=y_0+a_1t, \tag{2}$$

其中 $t=0,\pm1,\pm2,\cdots$.

注 我们把(2)称为(1)的通解公式.

证 证明分两步:1)(2)是(1)的解;2)(1)的任一解都可表示为(2)的形式.

1)证明的办法是,代入验算,既然 x_0,y_0 是(1)的解,当然满足(1),即 $ax_0+by_0=c$,由此

$$\begin{aligned}a(x_0-b_1t)+b(y_0+a_1t)&=c+(ba_1-ab_1)t\\&=c+(b_1da_1-a_1db_1)=c.\end{aligned}$$

这表明(2)式是(1)的解.

2)只需证明,(1)的任意解都具有(2)的形式.设 x',y' 是(1)的任一解,则 $ax'+by'=c$,从此减去 $ax_0+by_0=c$,即得

$$\begin{aligned}&a(x'-x_0)+b(y'-y_0)=0\\ \Rightarrow\ &a_1d(x'-x_0)+b_1d(y'-y_0)=0.\end{aligned}$$

消去 d,并移项,得到

$$a_1(x'-x_0)=-b_1(y'-y_0).$$

因为 $(a_1,b_1)=1$,所以 $a_1\mid(y'-y_0)$,从而

$$y'-y_0=a_1t\ \Rightarrow\ y'=y_0+a_1t.$$

将 $y'-y_0$ 代入上式即得

$$\begin{aligned}a_1(x'-x_0)=-b_1a_1t\ \Rightarrow\ x'-x_0\\=-b_1t\ \Rightarrow\ x'=x_0-b_1t.\end{aligned}$$

因此 x',y' 可表成(2)的形状.

综合1),2)知,(2)表示(1)的一切整数解.证毕.

例 求 $10x-7y=17$ 的全部解.

解 由观察法可知 $x=1,y=-1$ 是一个特解,因此一般解为

$$\begin{cases} x = 1 + 7t \\ y = -1 + 10t \end{cases} \qquad t = 0, \pm 1, \pm 2, \cdots.$$

5.1.2　整数的模

为了给出二元一次不定方程可解的充要条件，我们需要整数的模的概念．当然证明二元一次不定方程可解的充要条件还有别的方法，但这种方法较为简洁．

定义　一个整数集合构成一个模，如果它对加减是自封的．

例　全体整数集构成一个模．因为任何两个整数的和与差仍是整数．

自然数集不构成一个模．因为两个自然数的差可能不再是自然数．如 $5 - 7 = -2$，不再是自然数．

例　任一自然数的倍数构成一个模．如 $\{3k\}$，$k = 0, \pm 1, \pm 2, \cdots$ 构成一个模．集合中只有含有 0 的模称为 0 模．

定理 2　1）任何模中必含有 0；

2）若 a, b 在模中，则 $ma + nb$ 也在模中，其中 m, n 为任意整数．

证　1）在模中任取一数 a，依模的定义，$a - a = 0$ 在模中．

2）若 a 在模中，则 $2a = a + a$ 在模中，$3a = 2a + a$ 在模中．从而，对任意 m，ma 在模中．同样 nb 也在模中，进而 $ma + nb$ 在模中，证毕．

定理 3　对任何两个整数 a 和 b，所有形如 $am + bn$ 的全体整数形成一个模．

此定理很明显，无需证明．

定理 4　任何一个非 0 的模都是某正整数倍数所组成的集合．

证　这个模中一定有一个最小的正整数，设这个正整数为 d．我们来证明，模中的所有其它数都是 d 的倍数，只要证明了这一点，

定理的证明就完成了.

今用反证法证之.如果不然,则模中一定存在一个数 n,它不是 d 的倍数.这时必有整数 q 和 r,使得

$$n = dq + r, 1 \leqslant r < d.$$

由模的定义,$r = n - dq$ 在模中,而 $r < d$,这与 d 的最小性相矛盾.

这就证明了,模中的任何数都是 d 的倍数.此外,由模的定义知,d 的倍数也在模中,所以这个模是由 d 的整数倍组成的集合.证毕.

定理 4 一下子把模的结构搞清楚了:它是某个整数的倍数的集合.这个整数一定是模中最小的正整数.我们应尽全力把它找出来.

有了这些准备之后,现在我们给出两个正整数 a 与 b 的最大公因数的另一定义,并证明它与通常的定义是一样的.

定义 设 a,b 是两个正整数.在定理 4 中取一切形如 $am+bn$ 的数所成之模,在定理 4 的证明中所得到的 d 称为 a 与 b 的最大公因数,记为 (a,b).

定理 5 (a,b) 具有如下性质:

1) 存在整数 x,y,使得 $(a,b) = ax + by$;

2) 对任意两个整数 x,y,必有 $(a,b) \mid ax + by$;

3) 如果任一整数 e,$e \mid a$,$e \mid b$,则 $e \mid (a,b)$.

注 1. 我们可把定理 5 中的 1) 叫表示定理,a 与 b 的最大公因数 (a,b) 可通过 a 与 b 表示出来.

2. 3) 指出 (a,b) 就是通常的最大公因数.

3. 定理 5 的证明并不困难.但这里用的方法在初等数学中是很少见到,它是从模的结构中推出的.这种方法值得学习.

证 1) (a,b) 在由 $am+bn$ 所组成的模中,所以一定存在整数 x,y,使 $(a,b) = ax + by$.

2) 由定理 4,模中的数都是 (a,b) 的倍数.

3) 因为由 1),存在 x,y 使得 $(a,b) = ax + by$,由此式可知,$e \mid a, e \mid b \Rightarrow e \mid (a,b)$.

注　定理 4 的证明已暗示了辗转相除法. 我们了举例说明之.

例　取 $a = 323$ 和 $b = 221$. 由辗转相除法可得

$$323 = 221 \cdot 1 + 102.$$

可见 102 在以 $a = 323$ 和 $b = 221$ 所形成的模中, 又

$$221 = 102 \cdot 2 + 17,$$

所以 17 也在这个模中. 因

$$102 = 17 \cdot 6,$$

所以 17 是该模中的最小正整数(记着 17 是素数)即 $17 = (323,221)$. 利用这个方法可以求出定理 5 1) 中的 x 和 y:

$$17 = 221 - 2 \cdot 102 = 221 - 2(323 - 221) = 3 \cdot 221 - 2 \cdot 323.$$

所以 $x = -2, y = 3$.

5.1.3　可解的充要条件

现在我们来研究二元一次方程可解的条件. 我们有

定理 6　二元一次方程

$$ax + by = c \tag{3}$$

有解的充要条件是 $(a,b) \mid c$.

证　"⇒". 假定(3)式有整数解, 设为 x_0, y_0, 则

$$ax_0 + by_0 = c.$$

但 $(a,b) \mid a, (a,b) \mid b \ \Rightarrow\ (a,b) \mid c$, 故条件的必要性得证.

"⇐". 若 $(a,b) \mid c$, 则 $c = c_1(a,b)$, 其中 c_1 是整数. 由定理 5, 存在两个整数 s, t 满足方程

$$as + bt = (a,b) \ \Rightarrow\ asc_1 + btc_1 = (a,b)c_1 = c.$$

令 $x_0 = sc_1, y_0 = tc_1$, 即得 $ax_0 + by_0 = c$, 所以(1)式有整数解 x_0, y_0. 证毕.

例　判定下列二元一次方程是否可解:

1. $10x - 7y = 17$;　　　2. $117x + 21y = 38$;

3. $18x + 24y = 9$;　　　4. $107x + 37y = 25$.

解 1. 有解,因为$(10,7)=1\mid 17$;

2. 无解,因为$(117,21)=3\nmid 8$.

3. 无解,因为$(18,24)=6\nmid 9$;

4. 有解,因为$(107,37)=1\mid 25$.

5.1.4 如何求二元一次方程的解

求解的过程一般分为三步:

1) 判断方程(1)是否有解.先求出最大因数(a,b),并判断是否有$(a,b)\mid c$.若$(a,b)\nmid c$,则方程无解,就此停步.

2) 若$(a,b)\mid c$,则方程有解.设法求出一组特解,然后利用公式(2)求出通解.

3) 如果特解不易求,则用辗转相除法求解.

下面的例子是具体求法.

例 求$107x+37y=25$的一切整数解.

解 前面已经判断,这个方程有解.下面用辗转相除法去求它.由方程得

$$37y=25-107x \Rightarrow y=\frac{25-107x}{37}=-2x+\frac{25-33x}{37}.$$

令$y_1=(25-33x)/37$,则y_1应是一个整数,于是得到一个新的不定方程

$$37y_1+33x=25.$$

又 $$33x=25-37y_1 \Rightarrow x=\frac{25-37y_1}{33}=-y_1+\frac{25-4y_1}{33}. \quad (4)$$

仿照上面,令$x_1=(25-4y_1)/33$,又得到一个新的不定方程:

$$33x_1+4y_1=25.$$

又 $$4y_1=25-33x_1 \Rightarrow y_1=\frac{25-33x_1}{4}=6-8x+\frac{1-x_1}{4}. \quad (5)$$

令$y_2=(1-x_1)/4$,这就得到

$$x_1+4y_2=1. \quad (6)$$

由此不难看出,(6) 的一切解是

$$x_1 = 1 - 4t, y_2 = t(t = 0, \pm 1, \pm 2, \cdots).$$

把这个结果代入(5),得

$$y_1 = -2 + 33t(t = 0, \pm 1, \pm 2, \cdots).$$

这样一来,(5) 的一切解是

$$x_1 = 1 - 4t, y_1 = -2 + 33t(t = 0, \pm 1, \pm 2, \cdots).$$

把这个结果代入(4),得到(4) 的一切解是

$$y_1 = -2 + 33t, x = 3 - 37t(t = 0, \pm 1, \pm 2, \cdots).$$

再把它们代回原方程,得出不定方程的解为

$$x = 3 - 37t, y = -8 + 107t(t = 0, \pm 1, \pm 2, \cdots).$$

直接验算可知,它们的确是解.

从这个例子可以看得很清楚,解的过程就是对整个不定方程辗转相除,依次化为等价的不定方程,直到出现一个变量的系数为 ±1 为止.在上例中是 $x_1 + 4y_2 = 1$.这样的方程可以直接解出.然后再依次反推上去,就可求出原不定方程的解,为了减少运算次数,在作带余除法时,要取绝对值最小的余数.

如果不定方程无解,则使用这种算法时,到某一步就会立刻看出,今举一例.

例 求 $117x + 21y = 38$ 的解.

解 $21y = 38 - 117x \Rightarrow$

$$y = \frac{1}{21}(38 - 117x) = -6x + 2 + \frac{1}{21}(9x - 4).$$

令 $y_1 = (9x - 4)/21$, y_1 应是整数,于是

$$21y_1 = 9x - 4 \Rightarrow x = \frac{1}{9}(21y_1 + 4) = 2y_1 + \frac{1}{9}(3y_1 + 4).$$

再令 $x_1 = (3y_1 + 4)/9$, x_1 也应是整数,从而

$$9x_1 = 3y_1 + 4 \Rightarrow y_1 = \frac{1}{3}(9x_1 - 4) = 3x_1 - 1 - \frac{1}{3}.$$

最后一式表明，x_1,y_1 不可能同时为整数.所以不定方程无解.

这种方法还可用来解三元一次不定方程.

5.1.5 二元一次方程的非负解

有时问题只要求非负解或正解，例如在开头提出的百钱买百鸡的问题就属于这类问题.现在我们来研究方程什么时候有非负解.利用通解公式(2)，这可归结为确定参数 t 的值，使 x,y 均为非负或均为正数.当 a_1,b_2 异号时，不定方程(1)可有无穷多组非负解或正解.所以我们只要讨论 a_1,b_1 均为正的情形.

我们知道，方程(1)有解的充要条件是 $(a,b)\mid c$.设 $(a,b)=d$，则 $a=a_1d,b=b_1d,c=c_1d$，其中 a_1,b_1,c_1 都是整数.于是

$$ax+by=c \iff a_1x+b_1x=c_1,(a_1,b_1)=1. \tag{7}$$

这样一来，我们可以只考虑形如

$$ax+by=c,\qquad (a,b)=1$$

的方程了.

定理 7 设 a,b,c 都是正整数，$(a,b)=1$，则当 $c>ab-a-b$ 时，不定方程(7)有非负解.解数为 $\left[\frac{c}{ab}\right]$ 或 $\left[\frac{c}{ab}\right]+1$.当 $c=ab-a-b$ 时，不定方程(7)没有非负解.

证 由于 $(a,b)=1$，所以方程(7)必有解.设 x_0,y_0 是方程(7)的一组特解.由通解公式(2)，要解 x,y 是非负解，参数 t 必须满足

$$x_0-bt\geqslant 0,\qquad t\leqslant\frac{x_0}{b};$$

$$y_0+at\geqslant 0,\qquad t\geqslant-\frac{y_0}{a};$$

从而
$$-\frac{y_0}{a}\leqslant t\leqslant\frac{x_0}{b}.$$

由于 t 取整数数值，所以上式就是

$$-\left[\frac{y_0}{a}\right]\leqslant t\leqslant\left[\frac{x_0}{b}\right]. \tag{8}$$

因此，方程(7)的非负解的组数为

$$N=\left[\frac{y_0}{a}\right]+\left[\frac{x_0}{b}\right]+1. \tag{9}$$

($t=0$ 时也是一解)利用性质

$$[u]+[v]\leqslant[u+v]\leqslant[u]+[v]+1,\text{可知}$$

$$\left[\frac{y_0}{a}+\frac{x_0}{b}\right]\leqslant N\leqslant\left[\frac{y_0}{a}+\frac{x_0}{b}\right]+1. \tag{10}$$

上式等号中只有一个成立. 由于 x_0,y_0 是特解，所以

$$ax_0+by_0=c\ \Rightarrow\ \frac{x_0}{b}+\frac{y_0}{a}=\frac{c}{ab}\ \Rightarrow\ \left[\frac{x_0}{b}+\frac{y_0}{a}\right]=\left[\frac{c}{ab}\right].$$

由(10)，

$$N=\left[\frac{c}{ab}\right]\quad\text{或}\quad N=\left[\frac{c}{ab}\right]+1.$$

当 $c>ab-a-b$ 时，

$$\begin{aligned}1-\frac{1}{b}-\frac{1}{a}=\frac{ab-a-b}{ab}&<\frac{c}{ab}\\&=\frac{ax_0+by_0}{ab}=\frac{x_0}{b}+\frac{y_0}{a}\\&=\left[\frac{x_0}{b}\right]+\left\{\frac{x_0}{b}\right\}+\left[\frac{y_0}{a}\right]+\left\{\frac{y_0}{a}\right\}\\&\leqslant\left[\frac{x_0}{b}\right]+\frac{a-1}{a}+\left[\frac{y_0}{a}\right]+\frac{b-1}{b}.\end{aligned}$$

即

$$\begin{aligned}1-\frac{1}{a}-\frac{1}{b}&<\left[\frac{x_0}{b}\right]+\left[\frac{y_0}{a}\right]+1-\frac{1}{a}+1-\frac{1}{b}\\&\Leftrightarrow-1<\left[\frac{x_0}{b}\right]+\left[\frac{y_0}{a}\right].\end{aligned}$$

由(9)，

$$N=\left[\frac{y_0}{a}\right]+\left[\frac{x_0}{b}\right]+1>0.$$

这说明方程(7)必有非负解.

最后证明，当 $c = ab - a - b$ 时，不定方程(7)没有非负解.用反证法.

假定方程(7)有非负解 x_1, y_1，则

$$ax_1 + by_1 = c = ab - a - b \iff a(x_1 + 1) + b(y_1 + 1) = ab. \tag{11}$$

又 $$(a, b) = 1 \Rightarrow b \mid (x_1 + 1), a \mid (y_1 + 1),$$

$$x_1 \geqslant 0, y_1 \geqslant 0 \Rightarrow y_1 + 1 \geqslant a \geqslant 1, x_1 + 1 \geqslant b \geqslant 1.$$

由(11)，

$$ab = a(x_1 + 1) + b(y_1 + 1) \geqslant ab + ab = 2ab.$$

这是不可能的，这个矛盾说明，当 $c = ab - a - b$ 时，不定方程(7)没有非负解.

现在我们用上面的定理来解张丘建的百钱百鸡问题.

例 鸡翁一，值钱五，鸡母一，值钱三，鸡雏三，值钱一，百钱买百鸡，问鸡翁母雏各几何?

解 设用 x, y, z 分别代表鸡翁、鸡母、鸡雏的数目，就得到下面的方程：

$$5x + 3y + \frac{1}{3}z = 100,$$

$$x + y + z = 100.$$

消去 z，再化简，得到

$$7x + 4y = 100,$$

先求这个方程的非负解.容易看出来，$x = 0, y = 25$ 是一组特解.由通解公式，

$$x = 0 - 4t, \qquad y = 25 + 7t.$$

$$-3 = -\left[\frac{25}{7}\right] \leqslant t \leqslant \left[\frac{0}{4}\right] = 0.$$

令 $t = 0, -1, -2, -3$，共得到4组解：(0,25)，(4,18)，(8,11)，(12,4).因此买鸡的各种情况如下表所示：

x	0	4	8	12
y	25	18	11	4
z	75	78	81	84

5.1.6　多元一次不定方程

所谓多元一次不定方程，就是可以写成下列形式的方程：

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = N, \tag{12}$$

其中 $a_1, a_2, \cdots, a_n, N$ 都是整数，$n \geqslant 2$，并且不失掉一般性，我们可以假定 $a_1, a_2, \cdots, a_n$ 都不等于零. 现在首先证明

定理 8　(12) 式有整数解的充分与必要条件是

$$(a_1, a_2, \cdots, a_n) \mid N.$$

这里 $(a_1, a_2, \cdots, a_n)$ 表示 $a_1, a_2, \cdots, a_n$ 的最大公因数.

证　设 $(a_1, a_2, \cdots, a_n) = d$.

1) 若 (12) 式有解，即有 n 个整数 $x'_1, x'_2, \cdots, x'_n$ 满足等式

$$a_1x'_1 + a_2x'_2 + \cdots + a_nx'_n = N,$$

则 $d \mid (a_1x'_1 + a_2x'_2 + \cdots + a_nx'_n)$，即 $d \mid N$，这就证明了条件的必要性.

2) 若 $d \mid N$，我们要用数学归纳法证明 (12) 式有解. 当 $x = 2$ 时，由定理 6，(12) 式有解. 假定上述条件对 $n-1$ 元一次不定方程是充分的. 今证上述条件对 n 元一次不定方程也是充分的.

令 $d_2 = (a_1, a_2)$，则 $(d_2, a_3, a_4, \cdots, a_n) = d \mid N$. 由归纳法假定，方程

$$d_2t_2 + a_3x_3 + \cdots + a_nx_n = N$$

有解，设其一解为 $t'_2, x'_3, \cdots, x'_n$. 再考虑

$$a_1x_1 + a_2x_2 = d_2t'_2.$$

由定理 6，$(a_1, a_2) = d_2 \mid d_1$，上式有解，设其一解为 x'_1, x'_2，则

$$a_1x'_1 + a_2x'_2 + \cdots + a_nx'_n = d_2t'_2 + a_3x'_3 + \cdots + a_nx'_n = N.$$

故 $x'_1, x'_2, \cdots, x'_n$ 是(12)式的解.这就证明了条件的充分性.证毕.

这里需要指出的是,二元一次不定方程的通解中含有一个参数 t,三元一次不定方程就要含有两个参数了,因为 x_1, x_2, x_3 中只有两个确定了才有确定的解.同样道理可知,n 元一次不定方程的通解中含有 $n-1$ 个参数。

我们给一个三元一次不定方程的例子.仍用辗转相除法来解.

例 求 $15x_1 + 10x_2 + 6x_3 = 61$ 的全部解.

解 由于 $(15,10,6) = 1 \mid 61$,所以不定方程有解.注意到 x_3 的系数的绝对值最小,我们把原方程化为

$$x_3 = \frac{1}{6}(-15x_1 - 10x_2 + 61)$$

$$= -2x_1 - 2x_2 + 10 + \frac{1}{6}(-3x_1 + 2x_2 + 1).$$

令 $x_4 = \frac{1}{6}(-3x_1 + 2x_2 + 1)$,这是一个整数.由此解出 x_2 得到

$$x_2 = \frac{1}{2}(6x_4 + 3x_1 - 1) = 3x_4 + x_1 + \frac{1}{2}(x_1 - 1).$$

再令 $x_5 = \frac{1}{2}(x_1 - 1)$,这也是个整数.由此解出

$$x_1 = 1 + 2x_5.$$

取 x_5 作为参数,依次反推上去,就得到

$$x_2 = 3x_4 + x_1 + x_5 = 1 + 3x_4 + 3x_5,$$

$$x_3 = -2x_1 - 2x_2 + 10 + x_4 = 6 - 5x_4 - 10x_5.$$

再取 x_4 作为参数,这就得到了原方程的通解,其中含有两个参数 x_4, x_5,把解表示得更明确些,取 $s = x_4, t = x_5$,解可写为

$$x_1 = 1 + 2t,$$

$$x_2 = 1 + 3s + 3t,$$

$$x_3 = 6 - 5s - 10t, \quad s, t = 0, \pm 1, \pm 2, \cdots.$$

习　题

1. 解下列不定方程：

1) $15x + 25y = 100$；　　2) $306x + 360y = 630$；

3) $9x + 24y - 5z = 1000$.

2. 把 100 分成两份，使一份可被 7 整除，一份可被 11 整除.

§5.2 勾股定理

这一节对勾股定理作深入讨论，并为费马大定理的讨论作些必要的准备.我们从下面的问题开始.

5.2.1 问题

A. 设 x, y 分别是直角三角形的两直角边的长度，而 z 是斜边的长度.求 x, y, z 皆为整数的直角三角形，或求方程

$$x^2 + y^2 = z^2 \tag{A}$$

的所有整数解.

B. 设 x, y, z 是任意三角形的三条边的长度，当 x, y, z 是整数时，求面积是整数的全体三角形.

注 这是问题 A 的推广；取消了问题 A 中一个角为直角的限制.

C. 求方程

$$x^n + y^n = z^n (n > 2) \tag{B}$$

的全部整数解.

5.2.2 第一个重要定理 —— 勾股定理

中国古代已经知道，用边长为 3，4，5 的三角形去确定直角.埃及人知道用这个原理去构建他们的金字塔.古代巴比伦人也知道勾股

定理.但是首先给出合乎逻辑的证明的可能是毕达哥拉斯.因而这个定理在西方叫毕达哥拉斯定理.它是初等几何中最精彩、最著名、最有用的定理.它的重要意义表现在那些方面呢?

a) 它的证明是论证数学的发端;

b) 它是历史上第一个把数与形联系起来的定理,即它是第一个把几何与代数联系起来的定理;

c) 它导致了无理数的发现,引起了第一次数学危机,大大加深了人们对数的理解;

d) 勾股定理是历史上第一个给出了完全解答的不定方程.它引出了费马大定理;

e) 它是欧氏几何的基础定理,并有巨大的实用价值.

因为这个定理的重要与著名,所以研究的人特别多,或许在整个数学中还找不到另一个定理,其证明方法之多超过毕达哥拉斯定理.E.S.卢米斯在他的"毕达哥拉斯定理"一书的第二版中收集了这一定理的370种证明方法,并分了类.

定理1 (勾股定理)若直角三角形的三条边的长度分别为 a, b, c,其中 c 为斜边,则

$$a^2 + b^2 = c^2.$$

反过来,若三角形的三条边 a, b, c 满足

$$a^2 + b^2 = c^2,$$

则该三角形是直角三角形.

注 这个定理是欧几里得的《原本》第一卷的命题47.欧几里得给出了一个巧妙的证明,下面就是这个证明.证明分为两部分:1)正定理;2)逆定理.

正定理的证明 如图5-1所示,$\triangle ABC$ 是直角三角形,AB 和 AC 是直角边.以 AB, AC 和 BC 边为分别作正方形 $ABFG$, $ACKH$ 和 $BCED$.我们要证

$$AB^2 + AC^2 = BC^2.$$

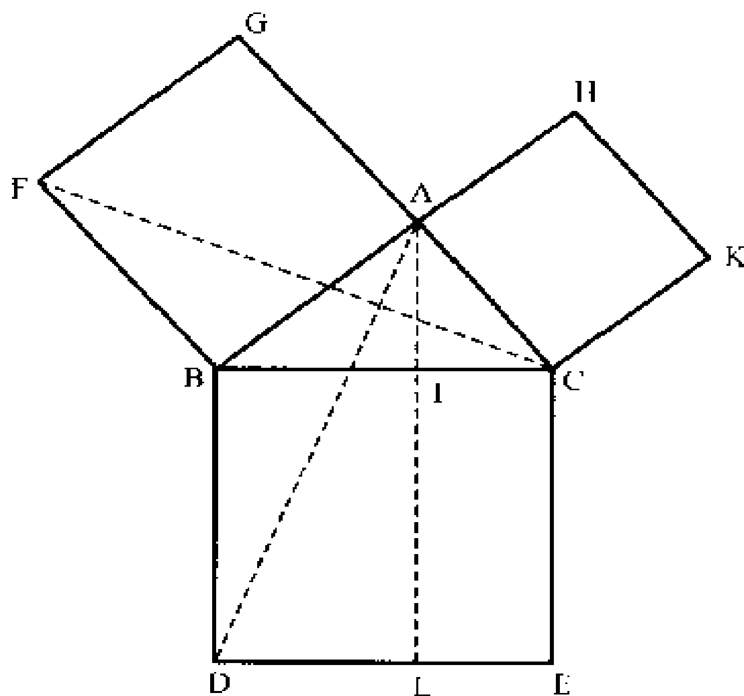


图 5－1

即要证，两直角边上的正方形面积之和等于斜边上正方形的面积. 作辅助线 AD，FC. 再作 $AL \perp DE$. 关键的问题是要证明 AC 和 AG 在同一条直线上. 因为 $\angle GAB$ 为直角，$\angle BAC$ 也是直角，所以这两个角的和等于两直角，这就证明了 GAC 是一条直线. 这是唯一一次应用了 $\angle BAC$ 是直角这一事实.

现在转向 $\triangle ABD$ 与 $\triangle FBC$. 这两个三角形的短边 AB 和 FB 相等，长边 BD 和 BC 也相等. 而

$$\angle ABD = \frac{\pi}{2} + \angle ABC, \angle FBC = \frac{\pi}{2} + \angle ABC,$$

所以 $\angle ABD = \angle FBC$. 根据三角形全等的边角边定理，$\triangle ABD \cong \triangle FBC$，因此这两个三角形的面积相等.

$\triangle ABD$ 与矩形 $BDLI$ 具有同一底边 BD，因此 $BDLI$ 的面积等于

$\triangle ABD$ 的面积的两倍．同样，$\triangle FBC$ 与正方形 $ABFG$ 也具有同一底边 BF．前面已经证明 GAC 是一条直线，因此正方形 $ABFG$ 的面积也是 $\triangle FBC$ 的面积的两倍．

综合这些结果，我们得出，矩形 $BDLI$ 的面积与正方形 $ABFG$ 的面积相等．

同样的方法可以证明，矩形 $CELI$ 的面积与正方形 $ACKH$ 的面积也相等．至此，勾股定理的证明已经在眼前了：

正方形 $BCED$ 的面积

= 矩形 $BDLI$ 的面积 + 矩形 $CELI$ 的面积

= 正方形 $ABFG$ 的面积 + 正方形 $ACKH$ 的面积．

这就完成了数学中的一个最重要的定理的证明．

用于证明定理的图形成了一个著名的图形．因为它的样子像风车，所以人们称它为“风车”．

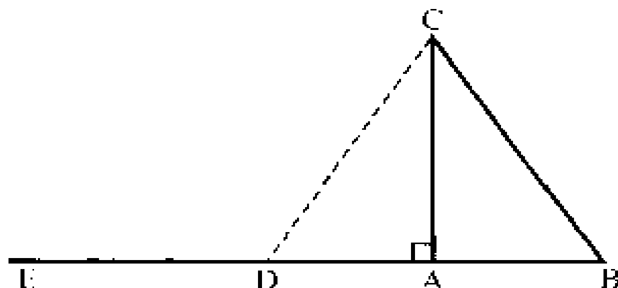


图 5 – 2

大多数人都知道勾股定理正定理的证明，但知道勾股定理逆定理证明的人比较少．欧几里得对逆定理的证明有两个特点，一个是证明简短，一个是证明逆定理时用到了正定理．这种逻辑方法没有前例．把逆定理建立在正定理的基础上，使这两个命题处于一种有序的地位．

逆定理的证明．如图 5 – 2，欧几里得先作出 $\triangle ABC$，并假定

$$BC^2 = AB^2 + AC^2.$$

下面证明 $\angle BAC$ 是直角．为此作 $AE \perp AC$，在 AE 上找点 D，使 $AD = AB$．连接 C 与 D，得到 $\triangle ADC$．根据正定理，

$$CD^2 = AD^2 + AC^2 = AB^2 + AC^2 = BC^2 \Rightarrow CD = BC.$$

根据全等三角形的"边边边"定理，$\triangle ADC \cong \triangle ABC$，从而

$$\angle CAB = \angle CAD = \frac{\pi}{2}.$$

这就完成了逆定理的证明．

正定理与逆定理结合在一起揭示了直角三角形的全部特征．

这个证明过去是，现在仍然是几何学中的最佳证明之一．将人们常见的直角与代数恒等式 $c^2 = a^2 + b^2$ 联系在一起，这是出乎意料的，正是这种联系，了这两个学科之间的相互联系，相互矛盾的运动，推动了数学的发展．

5.2.3 勾股定理的几何方面

勾股定理包括几何与数论两个方面．

几何方面指，直角三角形斜边的平方等于两直角边的平方和．边的平方的几何意义就是以该边为边的正方形的面积．因而勾股定理的几何方面表现为面积关系．

勾股定理的一些几何证明就来自这里．例如，图 5－3 四个全等三角形加一个正方形拼成一个大正方形，大正方形面积等于

$$(a+b)^2 = a^2 + 2ab + b^2.$$

它又等于

$$4 \times \frac{ab}{2} + c^2 = 2ab + c^2.$$

两者相等给出

$$a^2 + 2ab + b^2 = 2ab + c^2,$$

即

$$a^2 + b^2 = c^2.$$

其它证法还有不少，不再举例.

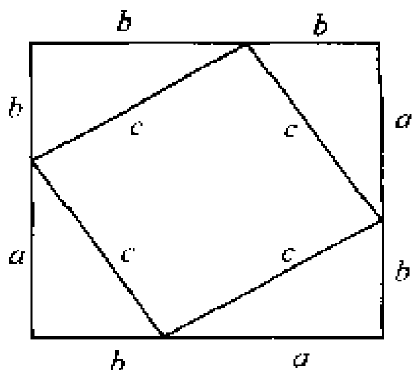


图 5－3

勾股定理中包含的面积关系还可以有更广泛的理解. 例如，把正方形改为以一边为直径的半圆，那么两直角边上的半圆面积之和等于斜边上半圆的面积.

5.2.4　勾股定理的数论方面

从数论方面看勾股定理就是求不定方程

$$x^2+y^2=z^2$$

的所有整数解，即本节一开始提出的问题 A.

化为数论问题后，已与原问题有了差别. 在几何问题中，x,y,z 是连续变量；在数论问题中，x,y,z 却是离散变量了，虽然要求严了，但在数论中却具有重意义.

问题 A 可以用整数的简单性质求解. 我们先对问题作些简化.

我们称满足方程 A 的整数组 (x,y,z) 为勾股弦三元数组，或毕达哥拉斯三元数组. 易见，对任意整数 p，(x,y,z) 与 (px,py,pz) 同时为勾股弦三元数组. 我们有

引理 1　若 (x,y,z) 是方程

$$x^2+y^2=z^2$$

的一组解，则 (px,py,pz) 也是方程的一组解，其中 p 是任意整数.

引理 1 所描述的性质称为方程的齐性，具有齐性的方程叫齐性方程. 例如，下面的方程都是齐性方程：

$$x+y=z;\ x^3+y^3+z^3=t^3;\ x^n+y^n=z^n.$$

有了引理 1，问题可简化为只考虑三个边没有公因数的三角形了，这样的三角形称为本原三角形，这样的数组 (x,y,z) 叫做本原三元数组. 例如 $(3,4,5)$ 就是一个本原三元数组，而 $(6,8,10)$，$(15,20,25)$ 是由 $(3,4,5)$ 生成的毕达哥拉斯三元数组，但它们不是本原的.

有些解一眼就能看出来，这就是 x,y,z 中一个或二个是零的解. 例如

$$0^2+y^2=z^2;x^2+0^2=z^2;x^2+y^2=0^2.$$

易见，$(0,-a,a)$，$(a,0,-a)$ 分别是第一、第二个方程的解，其中 a 是任何整数. $(0,0,0)$ 是第三个方程的解。这种有零的解称为平凡解，没有意思. 以后我们只考虑非平凡解，即 $xyz\neq 0$ 的解. 为了确定起见，不妨设 $x>0,y>0,z>0$.

在本原三元数组 (x,y,z) 中要求三个数没有公因数. 事实上，我们可以给出更强的定义：在本原三元数组中，任何两个数均无公因数.

引理 2 若 (x,y,z) 是本原解，则 $(x,y)=1$，$(x,z)=1$，$(y,z)=1$.

证 反证法. 假定不然，例如 x 与 y 有公因数 p. 我们设 $x=px_1$，$y=py_1$，则由 $z^2=p^2x_1^2+p^2y_1^2$ 可知，p 一定整除 z，这样一来，(x,y,z) 就不是本原解. 与假定矛盾，证毕.

引理 2 指出，x,y 不能同时为偶数. 这样必有一个为奇数，一个为偶数；或者皆为奇数. 我们再指出，两个直角边皆为奇数的情况也不存在. 为此，设

$$x=2a+1,y=2b+1,$$

那么

$$\begin{aligned}z^2=x^2+y^2&=(2a+1)^2+(2b+1)^2\\&=2+4a+4a^2+4b+4b^2\\&=2+4(a+a^2+b+b^2).\end{aligned}$$

这个数可被 2 整除，但不能被 4 整除. 这意味着 z^2 可被 2 整除，不可被 4 整除，这是不可能的. 因此，两个直角边一定是一奇、一偶.

因为 x 和 y 中一个是偶数，另一个是奇数，所以 z 一定是奇数. 事实上，设 $x=2a+1$，$y=2b$，则

$$z^2=x^2+y^2=(2a+1)^2+(2b)^2=4a^2+4a+4b^2+1.$$

由此知道，z 一定是奇数.

以下假定 x 是奇数，y 是偶数.

定理 2 满足方程 A 的本原解可表示为

$$x = m^2 - n^2, y = 2mn, z = m^2 + n^2, \tag{1}$$

其中 m, n 为正整数，满足条件

1) $(m, n) = 1$;

2) $m > n$;

3) m 与 n 中一个是偶数，另一个是奇数.

反过来，若 m 与 n 满足条件 1),2),3)，则(1) 是方程 A 的本原解.

下面给出定理 2 的三种证法：初等方法；几何方法；高斯整数法.

5.2.5 初等方法

首先指出，恒等式

$$(2mn)^2 + (m^2 - n^2)^2 = (m^2 + n^2)^2$$

保证了(1) 是方程 A 的本原解. 所以，为了证明定理，我们需要证明

1) 如果 x, y, z 是一组本原解，而且 x 是奇数，则它们一定具有(1) 的形式，其中 m, n 满足条件 1),2),3).

2) 如果 m, n 满足条件 1),2),3)，那么由(1) 给出的 x, y, z 是本原的.

证 1) 首先，由

$$x^2 + y^2 = z^2,$$

得

$$y^2 = z^2 - x^2 = (z + x)(z - x). \tag{2}$$

x, z 皆为奇数蕴含 $(z + x)$, $(z - x)$ 皆为偶数，所以 $(z + x)/2$, $(z - x)/2$ 都是整数，这样一来，

$$\left(\frac{y}{2}\right)^2 = \frac{z + x}{2} \cdot \frac{z - x}{2}. \tag{3}$$

令 $m_1 = \frac{1}{2}(z + x)$, $n_1 = \frac{1}{2}(z - x)$，那么

$$m_1 > n_1. \tag{4}$$

这时
$$\left(\frac{1}{2}y\right) = m_1 n_1. \tag{5}$$

其次，$(m_1, n_1) = 1$. 若 m_1 与 n_1 有公共素因数 p，则
$$p \mid m_1, p \mid n_1.$$
从而
$$p \mid (m_1 + n_1) = z, p \mid (m_1 - n_1) = x.$$
再由(3)，$p \mid y$，这与(x, y, z)的本原性相矛盾. 从而，$(m_1, n_1) = 1$.

由因子唯一分解定理，若 n^2 可以表示为两个互素整数 p, q 的乘积：$n^2 = pq$，则 p, q 也是完全平方数. 因此(5)中 m_1, n_1 也是完全平方数，即
$$m_1 = m^2, n_1 = n^2, (m, n) = 1. \tag{6}$$
不失一般性，可设 $m > 0, n > 0$. 在(4)，(5)中分别用 m^2, n^2 代替 m_1, n_1，我们就得到了勾股方程的本原解
$$x = m^2 - n^2; y = 2mn, z = m^2 + n^2. \tag{7}$$
(6)指出，$(m, n) = 1, m > n$ 保证了x, y, z都是正的. 此外，m 与 n 不可能都是奇数. 若 m, n 都是奇数，则由(7)知，y 与 z 将都是偶数，与(x, y, z)是本原数组相矛盾. 因此，m 与 n 满足条件 1)，2)，3).

2) 若$(m, n) = 1$，则$(m^2, n^2) = 1$. 若 m, n 中恰有一个是偶数，则 m^2, n^2 中也恰有一个是偶数，并且
$$x = m^2 - n^2, z = m^2 + n^2$$
都是奇数，若 x, z 有公因数，则公因数 p 必为奇数. 由此，数
$$x + z = 2m^2, z - x = 2n^2$$
也将以奇数 p 为公因数. 但这和$(m^2, n^2) = 1$相矛盾. 因此 x, z 没有公因数.

现在令 $y = 2mn$，就得到
$$x^2 + y^2 = (m^2 - n^2)^2 + 4m^2n^2 = (m^2 + n^2)^2 = z^2$$
由此可断定，x, y, z 是本原解. 因为如果 x 与 y 或 y 与 z 有公因数，那么 x 与 z 也会有公因数，与本原的定义相矛盾.

易见，m，n 不可能都是奇数，因为这时 x，z 将都是偶数．因而 m，n 中一个是奇数，一个是偶数．

这就证明了所有的本原解都是由(1)生成的，其中 m，n 满足条件 1)，2)，3)．

例如，取 $m = 11$，$n = 8$，那么 m，n 满足我们的条件，并得到

$$x = 57, y = 176, z = 185.$$

在下面的表中，我们给出了对应于最初的一些值 m 与 n 的所有本原三角形：

$n \backslash m$	2	3	4	5	6	7
1	3,4,5		15,8,17		35,12,37	
2		5,12,13		21,20,29		45,28,53
3			7,24,26			
4				9,40,41		33,56,65
5					11,60,61	
6						13,84,85

在结束这一节的时候，我们来叙述一个值得注意的性质：在本原三数组 x，y，z 中，x 与 y 中有一个能被 3 整除；x 与 y 中有一个能被 4 整除；x，y 与 z 中有一个能被 5 整除．例如三数组(8，15，17)中，15 能被 3 整除，8 能被 4 整除，15 能被 5 整除．我们把它作为习题留给读者．

5.2.6 几何方法

因为考虑非平凡解，所以我们将勾股方程

$$x^2 + y^2 = z^2 \tag{8}$$

化为

$$\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 = 1. \tag{9}$$

(8)的一组整数解对应于(9)的一组有理数解．反过来，从(9)的一组

有理数解，经过通分可以得出(8)的一组整数解．令 $u = x/z, v = y/z$，(9)化为

$$u^2 + v^2 = 1. \tag{10}$$

在几何上它表示单位圆．因而问题归结为求单位圆上的有理点问题．现在我们设法来求出这些有理点．

为此过点(1,0)作一直线 L，使之与单位圆相交．设直线的斜率为 k，由直线的点斜式方程知，直线 L 的方程为

$$v = k(u - 1). \tag{11}$$

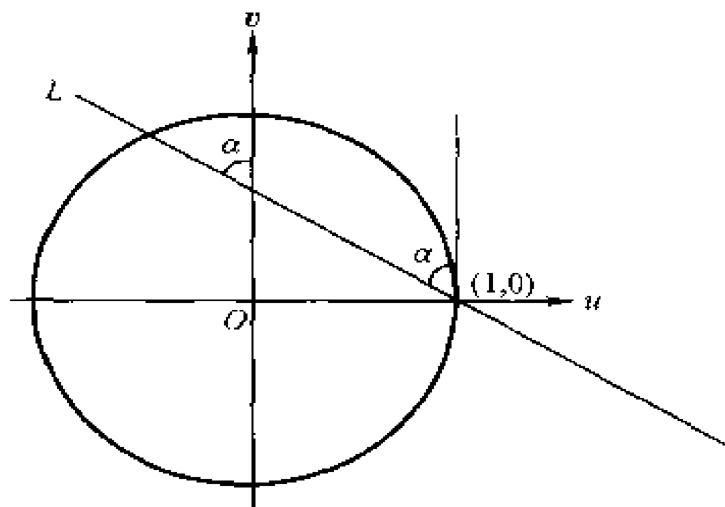


图 5－4

设直线 L 与 v 轴的夹角为 α(图 5－4)，则

$$k = \tan(\alpha + \frac{\pi}{2}) = -\cot\alpha.$$

代入(11)，得到

$$v = -(u - 1)\cot\alpha \Rightarrow \tan\alpha \cdot v + u = 1.$$

令 $t = \tan\alpha$，得到直线 L 的方程为

$$u + tv = 1. \tag{12}$$

当 α 从 0 变到 π 的时候，直线 L 扫过单位圆周上的每一点，相应地，t

将通过一切实数.但是我们知道,单位圆关于坐标轴是对称的,所以只需研究第一象限部分的单位圆就够了.这时 α 从 0 变到 $\frac{\pi}{4}$,从而 t 从 0 变到 1.

下面求直线(12)与圆(10)的交点.列方程

$$u + tv = 1,$$
$$u^2 + v^2 = 1.$$

易见,当 u, v 为有理数时,t 也为有理数.把直线方程代入单位圆的方程,得

$$(1 - tv)^2 + v^2 = 1,$$
$$1 - 2tv + t^2v^2 + v^2 = 1,$$
$$(1 + t^2)v^2 = 2tv.$$

当 $v \neq 0$ 时,有

$$v = \frac{2t}{1 + t^2}.$$

代入(12)得

$$u = \frac{1 - t^2}{1 + t^2}.$$

可令 $t = \frac{n}{m}, (m, n) = 1$,代入上式得出全部有理解:

$$u = \frac{m^2 - n^2}{m^2 + n^2}, v = \frac{2mn}{m^2 + n^2}.$$

对应于勾股方程的本原解

$$x = m^2 - n^2, y = 2mn, z = m^2 + n^2.$$

5.2.7 高斯的复整数

1796 年高斯证明了数论中一个非常重要的定理,这就是二次互反定律,用于解二次同余方程.当他试图将这一结果推广到高次同余方程时,发现使用形如 $a + bi$ 的数更为简洁,这里 a, b 是整数,i 是虚单位.这种形式的数叫高斯复整数.例如,下面的数都是高斯复整数:

$$2+3i, 2-3i, -1+4i, -2-i.$$

借助复整数可使我们对普通整数的认识加深.复整数是走向代数整数的第一步,并进一步发展为类域论、构成代数数论的核心内容,与费马大定理的解决密切相关.

高斯复整数是整数的最简单的推广,与普通整数最为接近.不难看出,两个高斯整数的和、差、积仍然是高斯整数,并且满足加法交换律、加法结合律、乘法交换律、乘法结合律,还有对加法与乘法满足分配律.这些性质与整数的性质是一样的,我们把具有这种性质的数集合叫作一个数环.所以整数构成一个环,高斯整数也构成一个环.

有关整数的最基本的问题是因子分解问题,对高斯整数也一样.要研究因子分解问题,就要考虑单位.对于自然数,单位就是1.对于整数,单位是1, -1.高斯整数的单位1, -1, i, -i.即高斯整数有四个单位数.

分解因子时,两数如果只差一个单位数的因子,我们称它们为相伴.对相伴的数我们不加区别.例如,$1+i, 1-i, -1+i, -1-i$这几个数是相伴的.$2+3i$与$2i-3$也是相伴的.但,$a+bi$与$a-bi$并不相伴.

高斯整数又与普通整数有明显的差异.数论的第一步是可除性理论.一个高期整数$a+ib$具有哪些因子?原来的素数还是不是素数?高斯的一个惊人发现是,许多原来的素数可以分解,也就是在高斯整数环中,它们不再是素数了.例如,

$$2=(1+i)(1-i), 5=(2+i)(2-i), 29=(5+2i)(5-2i).$$

而且他还发现,除2以外,可分解的素数都具有$4n+1$的形式.$4n+3$型的奇素数,如7,11,19等在高斯整数环中不能再分解,它们都是高斯素数.

现在我们来研究高斯素数.

现在的问题呢,除了$4n+3$型的素数是高斯素数之外,还有哪些其它类型的高斯素数呢?由2的分解因子可以知道,$1+i, 1-i$(还有

$-1+i$, $-1-i$）是高斯素数，其它还有那些整数是高斯素数呢？这里我们要用到复数的范数的概念. 一个复数 $a+ib$ 的范数定义为 a^2+b^2. 由于两个高斯素数 $a+bi$ 与 $c+di$ 乘积的范数等于范数的乘积，因此，要高斯整数是高斯素数，则它的范数应是素数或 $4n+3$ 型素数. 而这就是高斯证明的. 根据高斯这个定理，依范数的大小，高斯素数为

$1+i$	（范数为 2），
$1+2i, 2+i$	（范数为 5），
3	（范数为 3^2），
$3+2i, 2+3i$	（范数为 13），
$1+4i, 4+i$	（范数为 17），
$5+2i, 2+5i$	（范数为 29），
$6+i, 1+6i$	（范数为 37），
$5+4i, 4+5i$	（范数为 41），
7	（范数为 7^2）.

这样在理论上我们由范数就可以列举出所有的高斯素数，但由于素数的无限性，这实际上是做不到的.

有了单位和素数就可以考虑因子分解问题了. 高斯整数的因子分解要复杂一些，可分为四种情况.

（1）普通整数. 先做通常的素因数分解，然后把因子中的 2 和 $4n+1$ 型的素数分解为高斯素数. 例如，

$$\begin{aligned}30 &= 2\times 3\times 5\\ &= (1+i)(1-i)\times 3\times(2+i)(2-i).\end{aligned}$$

（2）纯虚数. 乘以 i 后变为普通整数，按（1）进行分解.

（3）普通整数或纯虚数乘以复高斯素数. 这时可以把倍数提出来，提出倍数后再用（1）. 例如

$$\begin{aligned}18+12i &= 6(3+2i) = 2\times 3(3+2i)\\ &= 3(1+i)(1-i)(3+2i).\end{aligned}$$

(4) 复整数(非素数). 这部分方法是新的,要用到复数的范数. 办法是先求出复整数的范数,这是一个普通整数,可进行因数分解. 它所对应的因子有三类:2,$4n+1$ 型素数和 $4n+3$ 型素数. 然后找出对应的复数.

例 求 $5+3\mathrm{i}$ 的因子.

解 $5+3\mathrm{i}$ 的范数为

$$5^2+3^2=25+9=34=2\times 17.$$

范数的平方为 2 的数有 $1+\mathrm{i}$ 或 $1-\mathrm{i}$. 范数为 17 的数有 $4+\mathrm{i}$ 或 $4-\mathrm{i}$. 简单计算指出

$$(1+\mathrm{i})(4-\mathrm{i})=5+3\mathrm{i}.$$

这里要注意的是,相同范数的复数不多,需一一验证.

在算术中最重要的定理是算术基本定理,即因子唯一分解定理:每一个整数可以唯一地分解为某些素数的乘积,最多相差 ±1. 高斯证明了,对于高斯整数,因子唯一分解定理也成立. 即,若不计因子的次序、单位数的因子,则素因数的个数和重数是唯一的. 因此,普通整数的那些整除性定理对高斯整数也成立. 这是一个极其重要的定理.

特别地,若 $z^2=pq$,$(p,q)=1$,则 p、q 是平方数.

5.2.8 类数问题

高斯整数除了在高次互反定律的研究上很有用以外,在别的问题上也很有用,其中包括与费马大定理的联系. 在各种数系中最富有成果的是形如 $a+b\sqrt{-d}$ 的数所构成的数系,这里 d 是自然数.

哪些 d 的值可以得到一种合理的数系呢(这里合理是指,这样的数系与整数系有类似的性质,特别是具有唯一因子分解定理)? 当 d 被 4 除余 3 时(记为 $d\equiv 3(\mathrm{mod}\,4)$),$a$,$b$ 既可取整数,也可取半整数. 例如,对 $d=3$,

$$\frac{1}{2}+2\sqrt{-3},\ \frac{3}{2}+\frac{5}{2}\sqrt{-3}$$

都是数系中的数. 当 $d\not\equiv 3(\mathrm{mod}\,4)$ 时,a,b 只取整数.

作了这一修正之后，我们看哪些 d 得到的数系具有唯一因子分解定理．前面已经指出，$d=1$ 时是高斯整数，有此定理．$d=2$，$d=3$ 时也有这个定理．但当 $d=5$ 时，唯一因子分解定理就不存在了．在这个数系中的 6 有两种因子分解式：

$$6=2\cdot3=(1+\sqrt{-5})(1-\sqrt{-5}).$$

再如，9 也有两种分解式：

$$9=3\cdot3=(2+\sqrt{-5})(2-\sqrt{-5}).$$

在高斯时代已经知道有 9 个 d 的值使 $a+b\sqrt{-d}$ 产生的数系有唯一因子分解定理．这 9 个值是 1，2，3，7，11，19，43，67，163．除此之外，还有别的值吗？没有了．1952 年瑞士数学家希格内尔（Kurt Heegner）证明了不存在第 10 个 d 的值．这就是说，除去这 9 个 d 的值外，由其它 d 得到的数系唯一因子分解定理都不成立．

下面的重要概念又是高斯引进的．高斯把每一个从 d 得到的数系与一个自然数 $h(d)$ 联系起来，并把这个自然数叫做那个数系的类数，用它来描述唯一因子分解定理失效的程度．如果 $h(d)=1$，则唯一因子分解定理成立．如果 $h(d)=2$，则唯一因子分解定理失效．例如 $d=5,6,10,13$ 时，$h(d)=2$．当类数为 3 时，唯一因子分解定理失效的程度更大．$d=23,31,59$ 等时，$h(d)=3$．类数越大，在该数系中，把数分解为素因子的方法就越多．

高斯注意到，每个类数 k，似乎有一个满足 $h(d)=k$ 的最大的 d 值．前面希格内尔的工作指出，使 $h(d)=1$ 的最大 d 值是 163．另外，使 $h(d)=2$ 的最大 d 值是 427，使 $h(d)=3$ 的最大 d 值是 907．高斯虽然猜到了这样一个结果，可是他既不能肯定，也不能否定．这样类数问题就是要对每个类数 k，确定使 $h(d)=k$ 的最大 d 值是否存在．经过 183 年的努力，在 1983 年由蔡革尔（Zagier）和格罗斯（Gross）所解决．

5.2.9 高斯复整数法

今用高斯整数来证明定理 2．由于

$$z^2 = x^2 + y^2 = (x + \mathrm{i}y)(x - \mathrm{i}y),$$

根据高斯整数的因子唯一分解定理和 $x + \mathrm{i}y$ 与 $x - \mathrm{i}y$ 互素，所以

$$x + \mathrm{i}y = \varepsilon(a + \mathrm{i}b)^2 = \varepsilon(a^2 - b^2 + 2\mathrm{i}ab), (\varepsilon = \pm 1, \pm \mathrm{i}).$$

不妨设 $\varepsilon = 1$，比较实部和虚部，得

$$x = a^2 - b^2, y = 2ab, z = a^2 + b^2.$$

证毕（当 $\varepsilon = \mathrm{i}$ 时，交换 x, y 的次序，这在本质上是一样的）.

§5.3　与勾股定理有关的问题

我们已经解决了毕达哥拉斯三角形的问题. 在数学中几乎总是这样，一个问题的解决就导致另外一些问题的解决，而新的问题常常可能比原来的问题困难得多. 这里也是如此.

与本原三角形有关的一个自然的问题是：当直角三角形中的一边已经给出时，如何去求另外两边的情形.

5.3.1　已知 x 边求本原三角形

根据 §5.2 的(1)式

$$x = m^2 - n^2 = (m + n)(m - n), \tag{1}$$

其中 m 与 n 满足 §5.2 定理 2 中的三个条件. 式(1)中的两个因数 $(m + n)$ 与 $(m - n)$ 是互素的. 为了证明这一点，我们注意到，由于 m, n 中一个是奇数，一个是偶数，所以

$$a = m + n, b = m - n \tag{2}$$

都是奇数. 若 a 与 b 有一个公共的奇素因数 p，那么 p 应该同时整除

$$a + b = m + n + (m - n) = 2m$$

及

$$a - b = m + n - (m - n) = 2n,$$

所以 p 应该同时整除 m 与 n. 但因 $(m, n) = 1$，这是不可能的. 于是，我们有

$$x = ab, a > b, (a, b) = 1. \tag{3}$$

这样一来，从 a, b 出发，令

$$m = \frac{1}{2}(a+b), n = \frac{1}{2}(a-b), \tag{4}$$

则 m,n 满足 §5.2 定理 2 的三个条件:1) $m > n$,2) $(m,n) = 1$,因为 m,n 的任一公因数必将整除 $a = m+n$ 及 $b = m-n$.3) m 与 n 不能都是奇数,否则,a 与 b 将均可能被 2 整除.这就证明了 m 与 n 满足定理 2 中的三个条件.有了这一结果,就可从 x 的分解式出发去定出本原三角形了.

首先将 x 分解为两个互素的奇数相乘:$x = ab$.然后由(4)定出 m,n.最后由 m,n 定出 y,z:

$$y = 2mn, z = m^2 + n^2.$$

例　设 $x = 15$.我们有两个形如(3)的分解式,即

$$x = 15 \cdot 1 = 5 \cdot 3.$$

由第一个分解式给出

$$m = 8, n = 7, x = 15, y = 112, z = 113;$$

再由第二个分解式给出

$$m = 4, n = 1, x = 15, y = 8, z = 17.$$

由例可以看出,将 x 分解为两个奇数相乘时,分解式可能有多种,因而答案也可有多种.

5.3.2　已知 y 边求本原三角形

设 y 边为已知.因为 m 与 n 中必有一个可被 2 整除,所以从 $y = 2mn$ 可看出,y 必须被 4 整除.把 $y/2$ 分解为两个互素的因数的乘积,并使其中一个是奇数,一个是偶数,就可把较大的一个取作 m,较小的一个取为 n.

例　设 $y = = 24$.我们有

$$\frac{1}{2}y = 12 \cdot 1 = 4 \cdot 3.$$

由第一个分解式给出

$$m = 12, n = 1, x = 143, y = 24, z = 145;$$

而由第二个分解式给出

$$m = 4, n = 3, x = 7, y = 24, z = 25.$$

5.3.3 已知 z 边求本原三角形

这是最后一种情形，将引导我们去触及数论中的一些重要问题. 如果 z 是本原毕达哥拉斯三角形的斜边，那么，根据 §2(1) 有

$$z = m^2 + n^2,$$

即 z 应该是满足条件 §2 定理 2 的两个数 m 与 n 的平方和.

这就导致我们去提出一个已被费马所解决的问题：什么时候一个整数 z 可表为两个平方数之和

$$z = a^2 + b^2?$$

暂时我们对 a 与 b 不加任何限制，它们可以有公因数，以及它们中的一个或全部可以是零. 在不超过 10 的整数中，以下的几个数是两个平方数之和：

$$0 = 0^2 + 0^2,\quad 1 = 1^2 + 0^2,\quad 2 = 1^2 + 1^2,\quad 4 = 2^2 + 0^2,$$
$$5 = 2^2 + 1^2,\quad 8 = 2^2 + 2^2,\quad 9 = 3^2 + 0^2,\quad 10 = 3^2 + 1^2.$$

其余的数 3,6,7 不能表为两个平方数之和.

下面我们来叙述：怎样判定一个数是不是两个平方数之和. 遗憾的是，它的证明不是简单的，这里只能略去不讲.

首先，我们来讨论素数. 每一个形如 $p = 4n + 1$ 的素数，一定是两个平方数之和. 例如，

$$5 = 2^2 + 1^2,\quad 13 = 3^2 + 2^2,$$
$$17 = 4^2 + 1^2,\quad 29 = 5^2 + 2^2.$$

一个令人惊异的事实是：这种表示式中唯一的.

其余的奇素数是 $q = 4n + 3$ 的形式. 如，

$$q = 3, 7, 11, 19, 23, 31, \cdots,$$

它们之中没有一个素数可表为两个平方数之和. 事实上，没有一个形如 $4n + 3$ 的数是两个平方数之和. 为了证明这一点，我们注意到：若

a 与 b 都是偶数，则 a^2 与 b^2 均可被 4 整除，所以 a^2+b^2 也被 4 整除；若 a 与 b 都是奇数，比如设

$$a=2k+1, b=2l+1,$$

那么
$$a^2+b^2=4k^2+4k+1+4l^2+4l+1$$
$$=4(k^2+l^2+k+l)+2,$$

所以 a^2+b^2 被 4 除后，余数为 1；最后，若整数 a,b 是一奇一偶，设 $a=2k+1, b=2l$，那么

$$a^2+b^2=4k^2+4k+1+4l^2.$$

所以 a^2+b^2 被 4 除后余数为 1. 因为这里列举出了 a 与 b 的所有可能性，所以我们证明了：两个平方数之和不可能是 $4n+3$ 的形式.

再注意到 $2=1^2+1^2$，这就完成了对所有素数的考察.

验证一个和数 z 是不是两个平方数之和，可按如下的方法进行：设 z 的素因数分解式是

$$z=p_1^{a_1}p_2^{a_2}\cdots.$$

那么，当且仅当每一个形为 $4n+3$ 的 p_i 的指数为偶数时，z 才是两个平方数之和.

例 $$z=198=2\cdot 3^2\cdot 11$$

不是两个平方数之和，因为 11 是 $4n+3$ 的形式的素数且是一次幂.

例 $$z=194=2\cdot 97$$

是两个平方数之和，因为它的两个素因数都不是 $4n+3$ 的形式，我们可求得

$$z=13^2+5^2.$$

让我们回到原来的问题上：确定所有的数 z，使它以作为本原毕达哥拉斯三角形的斜边. 这样的数 z 必须可以表示为 $z=m^2+n^2$，其中 m 与 n 满足 §5.2 定理 2 的条件.

例 1) $z=41$. 这里可求出唯一的一个把 z 表示为两个平方数之和的表达式

$$z=5^2+4^2.$$

所以

$$m=5, n=4, x=40, y=9, z=41$$

就是所对应的三角形.

2) $z=1105=5\cdot 13\cdot 17$. 我们有四个把 z 表示为两个数平方数之和的表达式

$$1105=33^2+4^2=32^2+9^2=31^2+12^2=24^2+23^2.$$

我们留给读者去求出相应的三角形.

关于毕达哥拉斯三角形的各种问题,都可以利用我们的公式

$$x=m^2-n^2, y=2mn, z=m^2+n^2$$

去解决. 例如,我们可以问,如何去求一个具有给定面积 A 的毕达哥达拉斯三角形. 如果这个三角形是本原的,那么它的面积是

$$A=\frac{1}{2}xy=mn(m-n)(m+n).$$

公式的四个因数中有三个是奇的. 不难看出,它们是两两互素的. 所以,为了求出所有可能的 m 与 n 的值,我们可以选挑选 A 的两个互素的奇因数 $k, l (k>l)$,并令

$$m+n=k, m-n=l.$$

这给出

$$m=\frac{1}{2}(k+l), n=\frac{1}{2}(k-l).$$

然后,再验证这些值是否确实满足 §2 定理 2 的三个条件,以确定有无这样的三角形.

例 求所有面积 $A=360$ 的毕达哥拉斯三角形. A 的素因数分解式是

$$A=2^3\cdot 3^2\cdot 5.$$

把 A 写为四个两两互素的因数的乘积的唯一方法是

$$A=8\cdot 1\cdot 5\cdot 9.$$

所以一定有 $m+n=9$. 这不能给出一个所需要的三角形:若 $m=8$,则 $n=1$,而 $m-n=7$ 不能整除 A. 另一种情形是 $n=8, m=1$,而这是为所需要满足的条件 $m>n$ 所排除的.

以上的结论，并未排除有一个非本原三角形具有 $A = 360$ 的可能性．这就是说，可以有一个非本原三角形，其面积为 360．若

$$dx, dy, dz$$

是边长具有公因数 d 的直角三角形的三边，那么，它的面积是

$$A = \frac{1}{2} \cdot dx \cdot dy \cdot = d^2 mn(m - n)(m + n).$$

所以 d^2 是 A 的因数，而且，如果 d 是三边长的最大公约数．那么

$$A_0 = \frac{A}{d^2} = mn(m - n)(m + n)$$

一定是一个本原三角形的面积．

我们来继续讨论．刚才讨论了 $A = 360$ 的情形，这个数有三个平方因数

$$d_1 = 4, d_2 = 9, d_3 = 36.$$

相应地可得

$$\frac{A}{d_1} = 90 = 2 \cdot 3^2 \cdot 5, \frac{A}{d_2} = 40 = 2^3 \cdot 5, \frac{A}{d_3} = 10 = 2 \cdot 5.$$

40 或 10 均不可能表为四个两两互素的因数的乘积，而对 90 只有一种这样的表法，即

$$90 = 1 \cdot 2 \cdot 3^2 \cdot 5 = 1 \cdot 2 \cdot 9 \cdot 5.$$

因为 9 是最大的因数，所以必须取 $m + n = 9$．从所有可能的选择 $m = 1,2,5$ 分别得到 $n = 8,7,4$，而仅有 $m = 5, n = 4$ 满足条件 $m > n$，但在这种情况下，$mn(m + n)(m - n) \neq 90$．所以，我们得到结论：没有一个毕达哥拉斯三角形，不管是本原的或不是本原的，其面积 $A = 360$．

我们还可以问许多其它的问题，但我们只再提出一个．一个三角形的周长是

$$c = x + y + z. \tag{5}$$

求出所有具有给定周长的毕达哥拉斯三角形．对于本原毕达哥拉斯

三角形,其周长是

$$c = 2mn + (m^2 - n^2) + (m^2 + n^2) = 2m(m + n).$$

我们把问题的解留给读者.请不要忘记,把你的方法应用于一些数值例子上.

我们已经解决了作出所有毕达哥拉斯三角形的问题.这引导我们去研究更一般的有关问题.一个自然的推广,是研究所谓赫伦三角形,这是以希腊亚历山大时期的数学家赫伦(Heron)的名字命名的.和以前一样,在这些三角形中,我们要求边长 x, y, z 是整数.但是我们放弃一个角是 $90°$ 的条件,而代之以要求面积是整数.显然毕达哥拉斯三角形属于这一类.

验证一个给定的三角形是不是赫伦三角形,最简单的办法是利用三角形面积的赫伦公式

$$A = \sqrt{\frac{1}{2}c\left(\frac{1}{2}c - x\right)\left(\frac{1}{2}c - y\right)\left(\frac{1}{2}c - z\right)},$$

这里 c 是我们在(5)式中定义的周长.虽然,我们知道很多很多的赫伦三角形,但是我们还没有一个给出它们全体的一般公式.这里是开头几个这种(非直角)三角形的例子:

$$x = 7,\ y = 15\ z = 20;$$
$$x = 9,\ y = 10\ z = 17;$$
$$x = 13,\ y = 14\ z = 15;$$
$$x = 39,\ y = 41\ z = 50.$$

习 题

1. 把上表扩充到所有 $m \leqslant 10$ 的值.

2. 求出斜边不超过 100 的所有本原三角形.

3. 在本原三角形中,证明 x 与 y 中有一个数能被 3 整除.(提示:如果 m, n 中有一个能被 3 整除,则 $y = 2mn$ 可被 3 整除.于是只要

证明,如果 m,n 都不能被3整除,则 $x=m^2-n^2$ 能被3整除.为了证明这一点,我们注意任意一个不能被3整除的整数 N,要么被3除余1,要么被3除余2,而这两种情况都有,N^2 被3除余1.)

4.在本原三角形中,证明,x,y,z 中有一个能被5整除(提示:借助上题的思路).

5.求出一边长等于22的所有直角三角形.

6.有没有面积为78,120的直角三角形.

7.若直角三角形的斜边长为1105,确定相应的本原三角形.

§5.4 费马大定理

前两节详尽地研究了毕达哥拉斯三角形,考虑了各种证法,研究了问题的各个侧面.这对我们有许多启发.正是这个问题启发了费马提出他的大定理.本节讨论费马大定理的主要解决过程.

5.4.1 费马和费马大定理

当费马在1665年去世的时候,他已经是欧洲最著名的数学家了.今天他被称为数论之父,但是在他那个时代,由于他在数论方面的工作是革命性的和超时代的,以致他的同时代人很少能理解.他的出名是因为他的其它贡献,其中包括解析几何.他与笛卡儿一起,分别独立地发明了解析几何;他是微积分的先导之一,又与帕斯卡一起奠定了概率论的基础.

费马作为数学家而著名,有两件事使人惊奇.第一,他是法学家,一生都在做法官和议员,数学只是他的业余爱好.第二,他生前从来没有发表过一篇作品.他的著作是在他死后他的儿子萨缪尔把他的文章,信件,以及对丢番图《算术》一书的批注等整理后发表的.

西兰德于1575年翻译出版了丢番图的《算术》的拉丁文译本,这是欧洲最早的译本,并作了很有价值的评注.法国人巴歇利用西兰德的译稿,于1621年出版了第一个拉丁文版本,并附有拉丁文的译文

和注释. 正是这个版本使费马走上了建立近代数论之路. 他在这个本子上作了许多批注,其中包括费马大定理. 费马的儿子萨缪尔将全部批注插入正文于 1670 年重新出版.

费马在读《算术》时,在有不定方程 $x^2+y^2=z^2$ 那页的边上,写出了具有历史意义的一段文字:

"但一个立方数不能分拆为两个立方数,一个四次方数不能分拆为两个四次方数. 一般说来,除平方之外,任何次幂都不能分拆为两个同次幂. 我发现了一个真正奇妙的证明,但书上的空白太小,写不下."

这就是说,费马已声称他证明了这一事实:不存在正整数 x, y, z 使

$$x^n+y^n=z^n, n>2.$$

这个命题称为费马大定理,或费马最后定理. 费马是否证明了这一定理呢?看来他像成千上万的后人一样,自以为证出来了,而实际上证错了.

自费马之后许多数学家花费巨大的劳动去解决这一问题. 经过 350 多年的努力,这问题终于由英国数学家维尔斯(Andrew Wiles 1953—)解决. 他的 108 页的论文《模曲线与费马大定理》于 1995 年 5 月在当代最有权威的数学杂志普林斯顿的《数学年刊》上发表了. 1996 年 3 月,维尔斯因此荣获沃尔夫奖.

费马大定理不仅是数论中的一个著名难题,更重要的是,正如著名数学家希尔伯特指出的,它是一只"会下金蛋的鹅". 它给整个数学带来了巨大财富,促进了代数数论和算术代数几何的建立,还发展了一系列先进的数学技术,形成了现代数论无尽的前沿.

我们把费马大定理证明过程中作出贡献的一些最突出的数学家罗列如下:

费马(Pierre de Fermat) 1601—1665

欧拉(Leonhard Euler)	1707—1783
热尔曼(Sophie Germain)	1776—1831
高斯(Carl Friedrich Gauss)	1777—1855
拉梅(Gabriel Lame)	1795—1870
狄里克雷(Peter Gustav Lejeune Drichlet)	1805—1859
刘维尔(Joseph Liouville)	1809—1882
库默尔(Ernst Eduard Kummer)	1810—1893
范迪维尔(Harry Schultz Vandiver)	1882—1973
志村五郎	1926—
谷山丰	1927—1958
符莱(Gerhard Frey)	
法尔廷斯(Gerd Faltings)	
吕贝特(Kenneth A. Ribet)	
维尔斯(Andrew J. Wiles)	

5.4.2　无穷递降法

费马发明了无穷递降法,并以此自豪.他晚年写了一封长信总结他在数论方面的发现.他很肯定地说,他的所有证明都使用了这一方法.这一方法是用来证明整数的某些性质或关系式不可能成立的.证明的办法是,如果这些性质或关系对任何正整数成立就会对某些更小的正整数成立.这样一来,用同样的推理方法可以证明,还有更小的正整数也成立.这种推理可以无限继续下去.但这是不可能的,因为一个正整数序列不可能无限地递减下去.下面我们用无穷递降法证明一个众所周知的定理.

定理1　不存在一对正整数 v,w 满足下列三条性质:

1) $(v,w)=1$;2) vw 是一个平方数;3) v,w 不都是平方数.

证　用反证法.假定存在满足上面三个条件的 v,w.即存在自

然数 u，使得

$$vw = u^2.$$

不妨设 v 不是平方数；否则可交换 v, w 的位置. 特别地 $v \neq 1$. v 至少有一个素因子，设它是 p，于是，$v = pk$. 由此

$$p \mid vw = u^2 \Rightarrow p \mid u,$$

所以 $$u = pm.$$

这时 $$vw = u^2 = (pm)^2 = p^2m^2.$$

又 $$vw = pkw = p^2m^2 \Rightarrow kw = pm^2 \Rightarrow p \mid kw.$$

由此可知 $$\Rightarrow p \mid k \text{ 或 } p \mid w,$$

但$(v, w) = 1$，所以 $p \mid k$. 设 $k = pv_1$，这时

$$kw = pv_1w = pm^2 \Rightarrow v_1w = m^2.$$

由 $v = pk = p^2v_1$ 推出，任何 v_1 的因数也是 v 的因数，所以$(v_1, w) = 1$. 另外，若 v_1 是平方数，则 $v = p^2v_1$ 也是平方数. v 不是平方数说明 v_1 也不是平方数. 这样一来，v_1, w 满足定理的三个条件，且 $v_1 < v$.

同样的推理指出，存在另一个正整数 $v_2 < v_1$，而 v_2, w 满足同样的三条性质. 无限次地重复这一推理，就可得出一个正整数的序列

$$v > v_1 > v_2 > \cdots.$$

这一序列是无穷递降的. 但这是不可能的，所以不存在满足三个条件的一对自然数 v, w. 证毕.

简言之，无穷递降法依赖于下述原理：要是一个给定的正整数满足一组给定的性质，就一定有一个更小的正整数满足同一组性质，从而不会有任何正整数满足这组性质.

定理 1 的等价叙述是

定理 2 若$(v, w) = 1$，且 vw 是平方数，则 v, w 一定都是平方数.

5.4.3 $n = 4$ 的费马定理

只要把无穷递降法与勾股定理的证明结合起来就可证明 $n = 4$

的费马定理.

假定存在 x, y, z,使得

$$x^4 + y^4 = z^4.$$

和勾股定理一样,可设 $(x, y) = 1, (x, z) = 1$. 因此, x^2, y^2, z^2 也是两两互素的,构成一组本原三元数组. 这样一来,我们有

$$x^2 = p^2 - q^2, y^2 = 2pq, z^2 = p^2 + q^2.$$

这里 $(p, q) = 1, p > q > 0, p, q$ 中一个奇数,一个偶数. 上面的第一个方程可以写成

$$x^2 + q^2 = p^2,$$

(x, q, p) 构成本原三元数组. 因此 p 是奇数, q 是偶数. 于是存在 a, b, 使得

$$x = a^2 - b^2, q = 2ab, p = a^2 + b^2,$$

这里 $(a, b) = 1, a > b > 0$. 这样一来,

$$y^2 = 2pq = 4ab(a^2 + b^2).$$

这就证明了 $ab(a^2 + b^2)$ 是一个平方数. 但是

$$(ab, a^2 + b^2) = 1,$$

这是因为任何素数 $p | ab$, 一定有 $p | a$ 或 $p | b$, 但不会既整除 a 又整除 b (a, b 互素), 因此 p 不会整除 $a^2 + b^2$. 这样一来, ab 和 $a^2 + b^2$ 都是平方数. ab 是平方数蕴含 a 和 b 都是平方数. 设 $a = X^2, b = Y^2$, 则

$$X^4 + Y^4 = a^2 + b^2$$

是平方数. 现在可以建立无穷递降序列了. 为此将前面结果总结一下. 我们从方程

$$x^4 + y^4 = z^4$$

出发,只用到 $x^4 + y^4$ 是平方数就能证明存在另一对正整数 X, Y, 使得 $X^4 + Y^4$ 也是一个平方数. 并且

$$X^4 + Y^4 = a^2 + b^2 = p < p^2 + q^2 = z^2 < z^4 = x^4 + y^4.$$

这样一来,我们就可以建立正整数的无穷递降序列了. 但是这种序列

不存在.因此两个四次幂的和不可能是一个平方数,更不会是一个四次幂.这就证明了 $n=4$ 的费马大定理.由此推出

系　对任何正整数 m,方程

$$x^{4m}+y^{4m}=z^{4m}$$

无解.

证　若方程有解 X,Y,Z,令 $X=x^m$,$Y=y^m$ 就得到 $x^4+y^4=z^4$ 的一组解.与前面的定理矛盾.

这样一来,费马定理对于4的倍数的 n 都成立.今考虑 $n>2$ 的任意指数.它或者可被大于2的素数整除,或者可被4整除,或者被两者整除.因而,在有了上面的系,要证费马大定理,只需证 n 是奇素数时,定理成立就行了.即只需证明,对任意奇素数 p,方程

$$x^p+y^p=z^p$$

无解即可.

更一般地,如果对于任意给定的指数 p 能证明费马大定理成立,那么定理对指数为 p 的任意倍数时也真.

5.4.4　$n=3$ 的情形

$n=3$ 的情形是欧拉证明的.在1753年8月4日给哥德巴赫的信中,他声称证明了 $n=3$ 时的费马大定理,但是没有给出证明.直到1770年,他才在彼得堡出版的《代数学引伦》给出了一个证明.这个证明有严重的缺陷.这一缺陷对于 $n=3$ 的情形尚可补救,而对于其它情形,类似的缺陷无法补救.

欧拉在证明中使用了形如 $a+b\sqrt{-3}$ 的数,其中 a,b 是整数.他指出,这种数形成一个数系.并且,进一步在这个数系中,他使用了唯一因子分解定理;他是从与整数类比得到的.前面曾指出,在形如 $a+b\sqrt{-d}$ 的数系中,只有9个整数根能保证唯一因子分解的性质.欧拉完全是靠运气才使他没有导致错误.

欧拉的证明有了突破,这是一个明显的进步.到20为止的 n 值,

只有6个值需要加以证明:5,7,11,13,17,19.但是,由于素数的个数是无限的,证明费马大定理仍是困难的事情.

5.4.5 初等方法的结束

欧拉注意到,$n=3$时的费马大定理的证明与$n=4$时的证明有很大的不相同,这就使人感到,证明费马大定理的一般情况还是非常遥远的事情.1816年人们重又燃起对费马大定理的兴趣,因为巴黎科学院为费马大定理设立了大奖和奖章.

1825年狄里赫勒和勒让德证明了$n=5$的情况.他们的方法基本上是欧拉对$n=3$时所使用的方法的延伸.在欧拉的证明中起关键作用的等式是

$$p+q\sqrt{-3}=(a+b\sqrt{-3})^3.$$

他们使用的类似等式是

$$p+q\sqrt{5}=(a+b\sqrt{5})^5.$$

他们没有使用唯一因子分解定理,因为这时它不成立.在寻找$n=7$的证明方法失败后,狄里赫勒在1832年证明了$n=14$的情况.拉梅在1839年证明了$n=7$的情况.但他必须使用某些与7结合得十分紧密的精巧工具.因此不采用新的方法,人们无望证明下一个$n=11$的情形.拉梅于1847年提出了新的证明思路.

拉梅证明思路的核心是利用n次单位复根.由此得出分圆整数.分圆整数也构成一个数系,在某种程度上类似于普通的整数.1847年3月1日,拉梅向巴黎科学院的成员作报告,声称他终于证明了费马大定理.他的关键思想是用了分圆整数.他犯了一个致命的错误,那就是他假定在分圆整数中存在唯一分解定理,但$a+b\sqrt{-3}$分圆整数并不存在这样的定理.刘维尔当场指出了这一错误,使拉梅非常窘迫.初等方法至此告一段落.

5.4.6 热尔曼的贡献

截止到1993年,数学家证明了,当n不超过4百万的时候费马定

理成立.你或许会惊叹地说,"啊,是够了!"的确,对工程师讲,这是足够了.但是 400 万与无穷比较,几乎是 0.

欧拉认为费马大定理是合理的.他猜想,一个平方数可以分解为两个平方数之和(这就是勾股定理),一个立方数至少要分解为三个立方数之和,一个四次方数不能分解为少于四个四次方数之和,如此等等.但有趣的是,在欧拉二百年后,兰德(Lander)和帕金(Parkin)利用计算机算出

$$144^5 = 27^5 + 84^5 + 110^5 + 113^5.$$

就在前几年,诺.埃尔基斯(Noam Elkies)在哈佛大学证明了,存在无穷多个情况,一个四次方数可以分解为三个四次方数之和.最小的情况是

$$422481^4 = 95800^4 + 217519^4 + 414560^4.$$

这说明欧拉的猜想是不对的.所以最杰出的人也有失足的时候.正像欧拉发现,费马数不都是素数一样.

但是像前面那样一个一个地证下去是没有穷尽的,数学家很清楚这一点.他们的目的当然是企图从个别中寻求一般规律.第一个得到一般定理的是法国女数学家索菲·热尔曼(Sophie Germain,1776—1831).当时女性在学术上受到歧视,所以她不得不用了一个男性的假名勒布郎(Leblanc)与一些大数学家通信,其中包括高斯和勒让德.

索菲·热尔曼幼年时在父亲的书房里发现了一些数学书,这些书深深迷住了她.但当她表示有志学习数学时,却遭到父母的反对.她只好偷偷地读.当她掌握了更多的数学后,就想学习更高级的数学.她想进入大学,而无法实现,只好在门外偷听.靠自学她掌握了微积分.之后她进入了巴黎综合工科学校的函授班(该校本科不收女生).热尔曼克服重重困难终于获得了成功.她对弹性片振动性质的透彻分析为她赢得了法兰西研究院奖金.她隐瞒了自己的身份,与世界上最优秀的数学家保持通信联系.1807 年高斯终于知道了她的真

实身份.她深感担忧,给高斯写了这样一封信:

"… 我以前曾用勒布朗的名字与您通信,这些信件无疑不值得您答复 … 我希望今天向您吐露的真情不会剥夺您给予我的荣幸,并恳请您抽出几分钟时间向我介绍一些您自己的情况."

高斯的回信充满慈爱与理解.他承认,当他看到勒布朗变成索菲·热尔曼的时候,确实感到吃惊,并对数学界中的不公正表示了自己的见解.高斯写道:

"我如何向您描述当我看到我的尊敬的信友勒布朗先生变为一个极为杰出的女士时是多么钦佩和吃惊呢?她给出了一个使人难以相信的光辉榜样.一般说来,对抽象科学,特别是对数的奥秘,很少人感兴趣.这门卓越的科学只向那些有勇气深入探索的人展现她迷人的魅力.由于我们的习惯和偏见,女性要熟悉这些棘手的研究必然遇到比男性多得多的困难.但是当一个女性成功地超越了这些障碍,深入到其中最难解的部分,那就毫无疑问,她必定具有最崇高的勇气,非凡的才能和超人一等的天才."

现在我们来谈热尔曼在费马大定理上的成就.为了解决困难问题,数学家总是对问题进行分解,分出容易的部分和困难的部分.先攻容易的部分再攻困难的部分.在费马大定理的证明史上对指数 n 有过两种划分.一种是分成第一种从属情形和第二种从属情形.一种是把素数指数 p 分为正则素数与非正则素数.第一种从属情形是指数 p 除不尽 x,y,z 中的任何一个.第二种情形是 $p|xyz$.热尔曼在1832年得到下述定理:

定理 如果 p 是奇素数,$2p+1$ 也是奇素数,则对指数 p,费马大定理的第一种情形成立.

这等于说,指数为 p 的费马方程有解,p 一定能整除该解的三个数中的一个.

如果 p 是素数,$2p+1$ 也是素数,则这种素数叫索菲·热尔曼素数.现在已知的最大的索菲·热尔曼素数是 2687145 · 3003 ·

$10^{5072}-1$,这是1995年10月由哈维·丢布纳(Harvey Dubner)发现的.但是,我们不知道是否有无穷多个这样的素数.

之后,热尔曼又得到了更一般的定理:

定理 设 n 是一个奇素数.若存在一个辅助素数 p,具有下述性质:

1) $x^n+y^n+z^n\equiv 0(\bmod p)\Rightarrow x\equiv 0$ or $y\equiv 0$ or $z\equiv 0(\bmod p)$;

2) $x^n\equiv n(\bmod p)$ 不可能成立,

则对 n 费马大定理第一种情形成立.

利用这个定理,热尔曼证明了,对小于100的素数,费马定理的第一种情况成立,勒让德把这一结果推广到所有小于197的素数,还有许多其它素数的情况.

热尔曼的一分为二虽然很有启发性,但是还不能对一批素数 p 证明费马大定理.

5.4.7 库默尔的工作和理想数

费马大定理的第一次真正突破来自德国的库默尔(Ernst Eduard Kummer 1810 - 1893).利用研究一个更重要的课题 — 高次互反定律 — 得出的思想,库默尔差一点解决了费马大定理.费马大定理中的数虽然都是通常的整数,但库默尔认为最好把它们看成"复数".这些复数由通常的整数与单位"虚根"构成,即前面提到的"分圆数域".

利用分圆数域,库默尔证明了指数是正则素数时的费马大定理.他找到了正则素数的判别法.在100以下的非正则素数只有37,59和67.1857年他对非正则素数改进了一个判据,对某一类非正则素数可以证明费马大定理.特别是对100以下的三个非正则素数37,59,67证明了费马大定理,但是在一个更难驯服的素数167前卡住了.遗憾的是库默尔的证明有严重错误.这个错误直到1926年才被美国数学家范狄维尔(H.S.Vandiver,1881—1975)纠正.

1929年范狄维尔得到对非正则素数费马大定理成立的判据.利

用这个判据，他把结果推进到 $p<211$. 以后逐年改进. 从 1954 年起主要通过计算机进行改进，而速度也大大加快. 如

1954 年，$p<2521$；1955 年，$p<4001$；1967 年，$p<25000$；

1975 年，$p<30000$；1976 年，$p<100000$；1977 年，$p<125000$；

1987 年，$p<150000$；1992 年，$p<1000000$；1994 年，$p<4000000$.

但这个办法不能彻底解决问题.

巴黎科学院在 1850 年又一次为解决费马大定理的人提供了一个金质奖章和 3000 法朗的奖金. 1856 年巴黎科学院收回成命，代之以给库默尔奖章，奖励他的"关于复数与单位根的结合而作出的漂亮研究". 在研究费马大定理的过程中，库默尔发展了代数数论.

直到现在我们还不能证明，是否有无穷多正则素数. 借助实验我们知道，素数中大约有 61% 的数是正则素数. 但这是物理，不是数学.

1908 年，波而·沃尔夫斯凯尔博士(Dr Paul Wolfskehl)在他的遗嘱中留下 10 万德国马克"给第一个证明费马大定理的人". 第一年这项奖金就吸引了 621 个申请者，这一下子使费马大定理大大出了名.

5.4.8　从丢番图到维尔斯

丢番图的"算术"究竟给了费马些什么呢?350 年来它带来何种影响呢? 丢番图的问题归结为解两个变量的方程，就是丢番图方程. 方程的图形是一条平面曲线. 丢番图希望找有理数的解. 50 年代以来，代数几何有了长足进步，对丢番图方程乃至更一般的数论问题提供了一个强有力的工具. 代数几何与解析几何的主要不同之处在于，解析几何用次数来对曲线和曲面进行分类，而代数几何则用亏格对代数曲线进行分类. 通过亏格 g 所有代数曲线可以分为三大类：

$g=0$：直线，圆，圆锥曲线；

$g=1$：椭圆曲线；

$g>1$：一般曲线，特别是费马曲线.

第一类是"有理曲线"包括阶为 1 和 2 的曲线. 在"有理曲线"上，

或不含有理点，或含无穷多个有理点．如果含无穷多个有理点，那么它们就能由一个简单的含有理函数的曲线给出．

例　在有理曲线 $x^2+y^2=1$ 上，所有的有理点由

$$x=\frac{1-t^2}{1+t^2},\ y=\frac{2t}{1+t^2}$$

给出，其中 t 是任一分数．

在 $x^2+y^2=-1$ 上没有有理点．

但在一般型曲线上，即在亏格大于等于 2 的曲线上，至多有有限个有理点．人们感兴趣的曲线是椭圆曲线，它们由 3 次方程来刻画．椭圆曲线不是椭圆，椭圆是圆锥曲线，而椭圆曲线是亏格为 $g=1$ 的曲线．

费马的问题实际上涉及许多方程．在

$$a^n+b^n=c^n$$

中令 $x=a/c,y=b/c$，可得

$$x^n+y^n=1.$$

$x=1,y=0$ 是显然解，不予考虑．当 $n=2$ 时，得到有理曲线，就是上面的例．而下面的两个方程

$$x^3+y^3=1,\ x^4+1=w^2$$

表示的曲线是不具有有理点的椭圆曲线．实际上，前一个方程是 $n=3$ 时的费马大定理；在后一方程中令

$$x=\frac{a}{b},\ w=\frac{c^2}{b^2}\Rightarrow a^4+b^4=c^4.$$

这正是 $n=4$ 时的费马大定理．

1983 年，法尔廷斯(Gerd Faltings) 证明了蒙德尔猜想：

法尔廷斯定理．亏格 $g\geqslant 2$ 的不可约代数曲线上只有有限多个有理点．

由此可以推出，如果 $p>3$，则

$$x^p+y^p=z^p$$

只有有限多个互素的整数解．这是费马大定理的第二次大突破．其证明需要来自代数几何的强有力的方法，只有专家才能看懂．这是一个重要成果．因此“纽约时报”在 1983 年 7 月 19 日对此作了报道．蒙德尔猜想的证明使数学家们看到，这将最终导致费马大定理的证明．

所以由此我们就知道了，在最坏的情况下，费马方程也只有有限的解．但是由库默尔引进的方法告诉我们，对许多 n，费马方程没有解，这启示给我们费马定理是真的．

还有一个关键的猜想必须提到，就是谷山 - 志村猜想．1955 年 9 月在东京召开了一次国际学术讨论会．在这个会上，日本的青年数学家谷山丰提出一些问题，并逐渐形成了一个重要的猜想．后来经志村五郎和魏尔精确成如下形式：有理数域上的每一条椭圆曲线都是模曲线．这一猜想称为谷山 - 志村猜想，或为谷山 - 志村 - 魏尔猜想．

从 60 年代初期开始，有人将费马方程和形如

$$y^2 = x(x + A)(x + B)$$

的椭圆曲线相联系．最初的着眼点是利用费马大定理的有关结论去证明与椭圆曲线有关的结论．1985 年德国数学家符莱（Gerhard Frey）迈出了关键的一步，从而成为费马大定理的第三次突破．符莱把研究方向倒转了过来．如果费马大定理有非零解（a, b, c），则可设计一条椭圆曲线

$$y = x(x + a^n)(x - b^n).$$

（它显然是有理数域上的椭圆曲线，我们称它为符莱曲线）．这条曲线不可能是模曲线．这个结论与谷山 - 志村猜想矛盾．如果符莱的结论和谷山 - 志村猜想都是正确的，那么“费马大定理不成立”就是错误的，由此得出费马大定理成立．1986 年，吕贝特（Ken Ribet）证明了符莱的论断．这样一来，证明费马大定理的工作归结为证明谷山 - 志村猜想．

得到这个消息之后，英国数学家维尔斯立刻集中全部精力去证明这个猜想．1986 年在巴黎的一个数论讨论班上，维尔斯声称，他将

证明费马大定理.七年之后,1993年6月23日,他在剑桥大学新成立的牛顿数学研究所作了以"模形式、椭圆曲线与伽罗华表示"为题的长达两个半小时的报告.在报告结束时他宣布:"我证明了费马大定理".这一下子轰动了整个数学界,并被誉为"世纪性的成就".但是事情不会如此顺利.同年11月15日他的老师柯兹指出,他的论文有漏洞.维尔斯很快承认,确有漏洞.但他认为他的证明路线没有错.他相信,他完全有能力解决这个问题.对此,他充满了信心.

1995年10月25日维尔斯发出两篇论文,一篇是他的"模椭圆曲线和费马大定理"(Modular Ellipticcurves and Fermat's Last Theorem),另一篇是与理查德.泰勒(Richard Taylor)合作的补篇"某些海克代数的环论性质"(Ring - theoritic properties of certain Hecke algebras).半年之后,1995年5月《数学年刊》(Annals of Mathematics 141930,May 1995)用一整期发表了这两篇论文.1996年3月维尔斯荣获了沃尔夫奖.费马大定理最终获得了证明.

5.4.9　费马大定理的推广

费马大定理经过350多年的努力,终于被攻克了.有些问题解决之后,就没有什么可干的了,费马大定理一类的不定方程可大不一样,它们解决之后,后面一大批问题接踵而来.这里给出两个有关的猜想.

费马 - 卡特兰猜想(The Fermat - Catalan Conjecture)　若 a, b, c 是互素的,则当 t, u, v 满足 $\frac{1}{t}+\frac{1}{u}+\frac{1}{v}<1$ 时,方程

$$a^t + b^u = c^v$$

只有有限个解.

1995年,达蒙(H.darmon)格兰维勒(A.Granville)找到10个解.前5个小解是

$$1+2^3=3^2, 2^5+7^2=3^4, 7^3+13^2=2^9,$$

$$2^7+17^3=71^2, 3^5+11^4=122^2.$$

5 个大解是

$$17^7 + 76271^3 = 21063928^2, 1414^3 + 2213459^2 = 65^7,$$
$$9262^3 + 15312283^2 = 113^7, 43^8 + 96222^3 = 30042907^2,$$
$$33^8 + 1549034^2 = 15613^3.$$

毕尔猜想(The Beal Conjecture) 对 A, B, C, x, y, z 是正整数，而 x，y，z 至少是 3，A, B, C 互素，方程

$$A^x + B^y = C^z$$

无解.

毕尔(Andrew Beal) 是一个银行职员，对数学很有兴趣. 他提出这一猜想，并且很慷慨，为猜想的解答者提供了 5 千美元的奖金，并且每年增加 5 千美元，最高到 5 万美元.

第六章 欧氏几何回顾

欧几里得的第五公设"也许是科学史上最重要的一句话".

C. J. Keyser

上个世纪最富有启发性和最值得注意的成就就是非欧几里得几何的发现.

D. 希尔伯特

我们大家从小就学习欧氏几何,知道两点决定一条直线,三点决定一张平面.我们都习惯性地认为我们所生活的空间是欧氏空间.我们对欧氏几何太熟悉了,或许会觉得欧氏几何还有什么可说的?其实不然,欧氏几何现在仍是中等教育中几何教育的核心.认清欧氏几何的地位和作用,它的主要成果是什么,它的缺陷是什么以及非欧几何的地位和作用,对几何教育的改革具有重要的意义.所以我们需要对欧氏几何作一认真回顾.

§6.1 欧几里得几何

几何学在希腊人手中成为数学的第一个分支,并趋于成熟.究其基本原因,几何乃是比较直观的数学形式,在日常生活中有直接应用.相反,代数在本质上更具抽象性,它包括一大堆符号,需要花费巨大的劳动才能掌握.

6.1.1 欧氏几何的诞生

欧几里得几何,简称欧氏几何,主要是以欧几里得平行公理为基础的几何学.公元前 7 世纪左右,希腊的著名数学家泰勒斯(公元前 625— 公元前 547 年)把埃及的数学知识传到希腊.泰勒斯又是第一

个在"知其然"的同时提出"知其所以然"的学者,而被公认为论证数学之父.他极力主张,对几何学的陈述不能凭直觉上的貌似合理就予以接受,相反,必需要经过严密的逻辑证明.他对几何学作出巨大贡献.他的主要贡献是,第一个证明了下列几何性质:

1) 对顶角相等.

2) 三角形内角和等于两直角之和.

3) 等腰三角形的两个底角相等.

4) 半圆上的圆周角是直角.

泰勒斯之后的另一位伟大数学家是毕达哥拉斯.他创立了毕达哥拉斯学派.他们研究了许多问题.五种正多面体,黄金分割,比例中项定理等,影响最大的有毕达哥拉斯定理及无理数的发现.雅典学派的希波克拉底,柏拉图,欧多克索斯提出几何三大问题:三等分任意角问题,立方倍积问题,化圆为方问题,对几何学的发展有很大贡献.柏拉图把逻辑思想引入几何学,使几何系统逐渐严格化.希腊人积累的几何知识同逻辑思想结合起来,为几何学的系统化,公理化奠定了基础,接着就是欧几里得的《几何原本》的出现.

约在公元前300年,在亚历山大城吸引的众多学者中有一位叫欧几里得,他来到亚历山大城创立了一所数学学校.欧几里得按照逻辑系统对几何学进行了整理,完成了数学史上最光辉的著作《几何原本》.这本书问世以后的两千年中一直被用作教科书,是学习几何知识和培养逻辑思维能力的典范教材.世界上大多数国家都有《几何原本》的译本.中国最古的译本是明代徐光启译出的,"几何"一词就是他第一个使用的.

与其说欧几里得创造了一种新数学,不如说他把旧数学变成一种清晰明确,有条不紊,逻辑严谨的新数学.这决非无足轻重的小事.必须认识到,《几何原本》绝不仅仅只是数学定理及其证明;早在泰勒斯时代,数学家已对命题作出过论证,而欧几里得对命题作了辉煌的公理化演绎,这是一个根本的区别.欧几里得成功地将零散的数学

理论编为一个从基本假定到最复杂结论的连续网络，所有这些都使之成为其后所有数学著作的范本.时至今日，在拓扑学、抽象代数、泛函分析等领域，数学家们还是首先提出公理，然后一步一步地推导，直至建立他们奇妙的理论.

欧氏几何的诞生是人类文明史上一个具有划时代意义的伟大事件.

6.1.2 《几何原本》的历史背景

欧几里得最著名的著作就是《几何原本》.尽管我们对古典时期了解甚少，但书中材料的主要来源一般都能查到.他的大部分材料来自柏拉图学派.据普罗克洛斯说，欧几里得把欧多克索斯的许多定理收入到《几何原本》中.他把前人只有马虎论证的结果变为无懈可击的证明.

对公理的选择，对定理的逻辑安排及一些定理的证明属于他.不过证明定理所采用的形式在奥托吕克斯的著作里已可看出，并且相当肯定地已为欧几里得以前的其他人所采用.尽管他从前人书里选取了许多材料，但欧几里得无疑是个大数学家.普罗克洛斯说过，希腊人对《几何原本》评价甚高，并引述了许多评语作为佐证.这些人中最重要的有赫伦(Heron)，波菲利(Porphyty，3 世纪) 帕普斯(Pappus，3 世纪末).由于《几何原本》写得如此之好，所以它才取代了其他人写的书.

6.1.3 欧氏几何的内容

《几何原本》共 13 卷，除其中第 5，第 7，第 8，第 9 和第 10 卷是讲授比例和算术理论外，其余各卷都是讲授几何内容的.第 1 卷包含平行线，三角形，平行四边形的定理；第 2 卷主要是毕达哥拉斯定理及其应用；第 3 卷讲授圆的定理；第 4 卷讨论圆的内接与外切多边形定理；第 6 卷的内容是相似的理论；最后 3 卷是立体几何.

《几何原本》是由定义，公设，公理组成的演绎推理体系.在第 1

卷开始他首先提出23个定义.前6个定义是

1）点没有大小.

2）线有长度而没有宽度.

3）线的界是点.

4）直线上的点是同样放置的.

5）面只有长度和宽度.

6）面的界是线.

我们列出欧几里得的公设和公理.以后几讲将要由此引出一些重要结论.

公设 1）给定两点,可连接一线段.

2）线段可无限延长.

3）给定中心和圆上一点可以作一圆.

4）所有直角彼此相等.

5）如一条直线与两条直线相交,并且在同侧所交出的两内角之和小于两个直角,则这两条直线无限延长后必在该侧相交.

公理 1）与同一个东西相等的东西,彼此也相等.

2）等量加等量,其和相等.

3）等量减等量,其差相等.

4）彼此重合的东西相等.

5）整体大于部分.

由此我们看到,前三个公设限定了用圆规和无刻度的直尺可以完成哪些作图.因此这两件仪器被称为欧几里得工具,使用它们可以完成的作图称为欧几里得作图.这种作图增加了几何学趣味.人们花费了大量的精力去解决几何三大难题,尽管是徒劳的,但从各方面推动了数学的发展.

在欧几里得几何体系中,第五公设与"在平面内过已知直线外一点,只有一条直线与已知直线平行"相等价.现在把后一命题叫作欧几里得平行公理.19世纪它导致了数学发展史上一些非常重要的结

果，这就是非欧几何的诞生.

6.1.4 欧氏几何的优缺点

《几何原本》是最早一本内容丰富的数学书，为所有的后代人所使用，它对数学发展的影响超过任何一本别的书. 读了这本书之后，对数学本身的看法，对证明的想法，对定理按逻辑顺序的排法，都会学到一些东西. 它的内容也决定了其后数学思想的发展. 因此，我们应该指出它有哪些特点如此深刻地影响了日后的数学.

这本书的陈述方式是史无前例的. 开头就列出所有公理和定义，然后有条不紊地推出一系列定理. 这是欧几里得所独创的. 虽然在他之前就有人提出，要先证明图形存在然后才能把它作为逻辑对象来处理，但是在他手里才终于把这一步准备工作作得很周密. 欧几里得对公理的选择搞得很出色. 他能用一小批公理证出几百个定理，其中有好多是深奥的. 他对平行公理的处理显得特别聪明. 他知道，这样的公理必然涉及到无限远空间，而这是人们经验之外的事. 他也认识到这样的公理不能省略. 于是就采取了这样一种办法，提出两条直线不能交于有限远处的条件. 并且他把无需平行公理的定理都放在了平行公理之前.

欧几里得虽然利用图形的重合来证明全等，但明显地他对这一方法是否完善有点不放心. 这个方法有两点值得怀疑：首先它使用了运动的概念，而这是没有逻辑依据的；其次重合法默认图形从一处移到另一处时所有性质保持不变，这就要对物理空间作许多假定. 欧几里得对这一方法不甚放心，其证据是，凡是能用其它方法证明的地方，他总不用这一方法，即使是重合法能给出更简单的证明.

虽然我们高度评价《几何原本》的内容在整体上的组织，但全书13篇并未哈成一气，在某种程度上是前人著作的堆砌. 例如第7,8,9篇对整数重复证明了先前对量所给出的许多结果. 第13篇的第一部分重复了第2,第4篇中的结果，等等.

直到19世纪的前半个多世纪，数学家一般都把欧几里得的著作

看成是严格性方面的典范,但也有少数数学家看出了其中的严重缺点,并设法纠正.19世纪末,几何领域中最敏锐的思想家日益关心《几何原本》缺乏真正的严密性问题.非欧几何的创立更加激发人们去探索古典几何的正确而又完备的叙述.

《几何原本》的主要缺陷是什么呢?首先,欧几里得的定义不能成为一种数学定义,有的不过是几何对象,如点,线,面等的一种直观描述,有的含混不清.这些定义在后面的论证中实际上是无用的.其次,欧几里得的公设和公理是远不够用的.因而在《几何原本》许多命题的论证中不得不借助直观,或者或明或暗地引用了用他的公设或公理无法证明的东西.例如,公设2)断定直线可被无限延长,但是它不一定意味着直线是无限长的,而只意味着,它是无端的,或无界的.连接球面上两点的大圆的弧可沿着大圆无限延长,但它不是无限长的.德国数学家黎曼在1854年所作的著名演讲《关于几何学基础的假定》中区别了直线的无界和无限长,成为黎曼几何诞生的起点.

19世纪末期,德国数学家D.希尔伯特于1889年发表了《几何基础》.书中成功地建立了欧几里得几何的完整的公理体系,这就是希尔伯特公理体系.他从叙述21条公理开始,其中涉及6个本原的或不定义的术语,即作为元素的点,直线和平面,以及它们之间的三种关系:"属于","介于"和"全等于".他把公理分为五类,分别处理关联,顺序,全等,平行和连续性.

关于引进不定义的术语,他曾作过一个著名的解释,强调保持它们完全抽象性的重要性.在任何演绎系统中,因为每个定理都是用前面定理证明的,前面定理的证明需要更前面的定理.这种倒推过程不能无限地进行下去,所以在开始必须作某些不证明的假设,这些假设称为公设或公理.它们是不是'真实的'或在什么意义下是'真实的',系统本身不能回答.公理系统的规定有任意性,只要它们不导致两个互相矛盾的定理.同样地,每个技术术语也必须参照前面的术语定义,所以在开始必须给出某些不定义的术语,它们是一些逻辑的概

念，它们在系统的应用中获得意义．当然，不定义的术语和公理的数目应尽可能地少．希尔伯特的几何体系是一种抽象的数学学说，现在看来，欧几里得几何只是它的一个模型而已．类此，用公理化方法可以建立高于三维的几何．

6.1.5　欧氏几何的历史地位

欧几里得的《几何原本》被称为数学家的圣经，在数学史，乃至人类科学史上具有无与伦比的崇高地位．它的主要贡献是什么呢？

1）成功地将零散的数学理论编为一个从基本假定到最复杂结论的整体结构．

2）对命题作了公理化演绎．从定义，公理，公设出发建立了几何学的逻辑体系，成为其后所有数学的范本．

3）几个世纪以来，已成为训练逻辑推理的最有力的教育手段．

6.1.6　几何学在数学教育中的地位

无论是中学还是大学的数学课程都发生过，并且正在发生着种种变革，其中最引人注目的是几何在课程中的核心地位的衰落．欧氏几何已从宝座上跌落下来．

在几个世纪里，欧几里得控制着数学舞台．但是代数的出现，笛卡儿将其应用于几何，以及随后微积分的发展，改变了数学的整个特征．数学变得更加符号化，更抽象了．

无可奈何花落去．

但是，也不要这样悲观，而应向事物的深层看去．英国著名数学家 M．阿蒂亚说了这样一段深刻的话，值得深思："几何是数学中这样一个部分，其中视觉思维占主导地位，而代数则是数学中有序思维占主导地位的部分．这种区分也许用另外一对词更好，即'洞察'与'严格'，两者在真正的数学研究中起着本质的作用"．这就明确指出，几何学不只是一个数学分支，而且是一种思维方式，它渗透到数学的所有分支．对这种思维方式应当给予足够的训练．

§6.2　尺规作图问题

6.2.1　几何三大难题

著名的几何作图三大问题是：

1）三等分任意角.

2）化圆为方：求作一正方形，使其面积等于一已知圆的面积.

3）立方倍积：求作一立方体，使其体积是已知立方体体积的两倍.

问题的难处在于限制用直尺和圆规.2000 多年来，数学家为解决三大问题投入了大量精力.如果解除这一限制，问题很容易解决.如化圆为方问题曾被欧洲文艺复兴时期的大师达·芬奇用一种巧妙的方法给出解答：取一圆柱，使其底和已知圆相等，高是半径的一半.将圆柱滚动一周，产生一个矩形，其面积为 $2\pi r \times r/2 = \pi r^2$.这正好是圆的面积.再将矩形化为正方形，问题就解决了.

那么用直尺和圆规能不能解决三大问题呢?答案是否定的，三大问题都是几何作图不能问题.答案不能只从几何本身去找.1637 年笛卡儿创立了解析几何，为解决尺规作图三大问题奠定了基础.1837 年，法国数学家旺策尔(Pierre L. Wantzel)证明了三等分任意角和立方倍积问题都是几何作图不能问题.化圆为方问题相当于用尺规作出 π 的值.1882 年法国数学家林得曼证明了 π 是超越数，从而证明了化圆为方的不可能性.

证明三大问题不可解的工具本质上不是几何的而是代数的.在代数还没有发展到一定水平时是不可能解决这些问题的.但是，正是在研究这些问题的过程中促进了数学的发展.两千多年来，三大几何难题引起了许多数学家的兴趣，对它们的深入研究不但给予希腊几何学以巨大影响，而且引出了大量的新发现.例如，许多二次曲线，三次曲线以及几种超越曲线的发现，后来又有关于有理数域，代数数与

超越数，群论等的发展.在化圆为方的研究中几乎从一开始就促进了穷竭法的发展，而穷竭法正是微积分的先导.

6.2.2 用尺规可作什么图

从中学几何里，我们知道下面的图是可作的：

1）二等分已知线段.

2）二等分已知角.

3）已知直线 L 和 L 外一点 P，过 P 作直线垂直 L.

4）任意给定自然数 n，作已知线段的 n 倍，以及 n 等分已知线段.

5）已知线段 a, b，可作 $a+b, a-b,, ab, a/b$，其做法如图（图 6－1）所示.接着 ra 也可作，这里 r 是正有理数. ra 这样作：设 $r = p/q$，p 和 q 都是自然数，因此 $ra = pa/q$.先作 a 的 p 倍 pa，再作 pa/q，这样 ra 就作出来了.

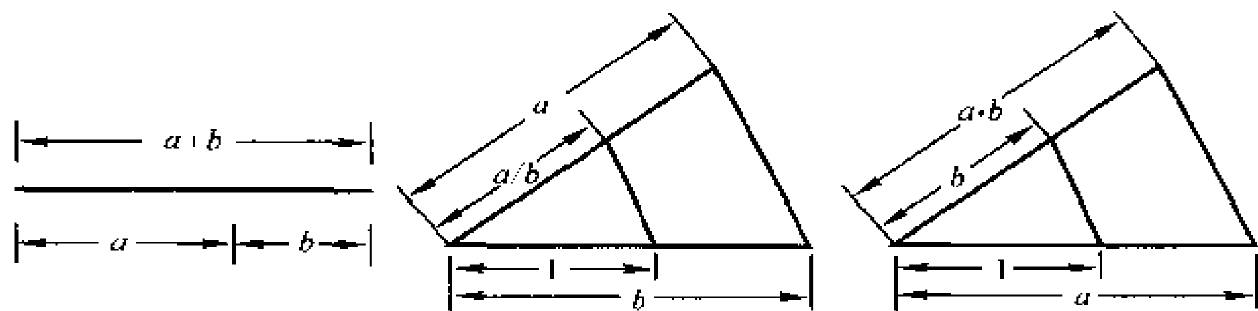


图 6－1

上面各条告诉我们，已知量的加，减，乘，除能用几何作图来实现.前面曾指出，从 0,1 出发利用算术运算可以构造出全部有理数，即构造出有理数域.现在我们知道了，只要给定单位 1，我们可以用尺规作出全部有理点.几何与代数在这里达到了完全的统一.

6）已知线段 a 作 $\sqrt{a}$.这一条超出了有理作图的范围.

如图 6－2，$OA=a$，$AB=1$．以 OB 为直径作圆．过 A 作 OB 的垂直线交圆周于 C．直角三角形 OAC 与直角三角形 OBC 有一个公共角 $\angle COB$，由此可得，$\angle OCA=\angle ABC$．这样一来，$\triangle OAC\sim\triangle ABC$．设 $AC=x$．我们有，

$$\frac{a}{x}=\frac{x}{1},x^2=a,\ x=\sqrt{a}.$$

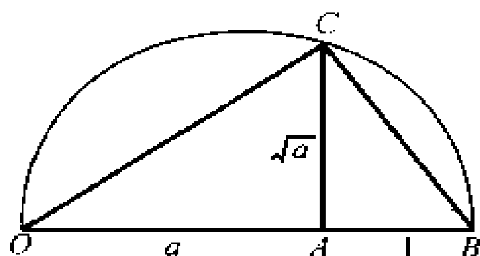


图 6－2

对尺规作图我们已有了初步知识，下面借助解析几何的理论将其深化．

6.2.3 有理数域的扩张．

我们需要对尺规作图作进一步分析．每一个直尺圆规作图都由一系列步骤组成，每一步都不外是下列做法之一：

1）用一条直线连接两点．

2）求两条直线的交点．

3）以一点为心，定长为半径作一圆．

4）求一个圆与一条直线的交点，或切点．

5）求两个圆的交点，或切点．

下面利用解析几何的知识对上面几条作进一步的分析．假定在平面上取好了直角坐标系，用 (x,y) 表示平面上的点．直线方程具有形式

$$ax+by+c=0, \tag{1}$$

圆的方程具有形式

$$x^2 + y^2 + 2\alpha x + 2\beta y + \gamma = 0. \tag{2}$$

系数 $a, b, c, \alpha, \beta, \gamma$ 都是有理数.

假定最初只给了一个元素 1，由 1 出发，我们能作出整个有理数域，从而能作出平面上的所有有理点，即两个坐标皆为有理数的点.我们能作出新的无理数，如$\sqrt{2}$，它不属于有理数域.从$\sqrt{2}$ 出发，通过"有理"作图，可以作出所有形如

$$a + b\sqrt{2} \tag{3}$$

的数，这里 a, b 是有理数.同样地，我们可以作出所有形如

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}}, \text{或} (a + b\sqrt{2})(c + d\sqrt{2})$$

的数，这里 a, b, c, d 是有理数.但这些数总可以写成(3) 的形式.例如

$$\begin{aligned}\frac{a + b\sqrt{2}}{c + d\sqrt{2}} &= \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \frac{c - d\sqrt{2}}{c - d\sqrt{2}} \\ &= \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2} = p + q\sqrt{2},\end{aligned}$$

这里 p, q 是有理数，且分母 $c^2 - 2d^2$ 不可能是零(为什么?).同样，

$$\begin{aligned}(a + b\sqrt{2})(c + d\sqrt{2}) &= (ac + 2bd) + (bc + ad)\sqrt{2} \\ &= r + s\sqrt{2},\end{aligned}$$

这里 r, s 是有理数.因此，由$\sqrt{2}$ 的作图，我们产生了全部形如(3) 的数集，其中 a, b 是任意有理数.由此得

命题 1　形如(3) 的数形成一个域.

证明　只需证，形如(3) 的两个数的和、差、积和商也是形如(3) 的数，且满足域的基本性质就行了，这是容易验证的.

这个域比有理数域大.事实上在(3) 中取 $b = 0$ 就可得到有理数域.有理数域是它的一部分，称为它的子域.但是，它显然小于全体实

数域.

将有理数域记为 Q,这个数域记为 Q_1,称它为 Q 的扩张. Q_1 中的数都可用直尺和圆规作出来.现在我们继续扩充可作数的范围.在 Q_1 中取一个数,如 $k = 1 + \sqrt{2}$.求它的平方根而得到可作图的数

$$\sqrt{k} = \sqrt{1 + \sqrt{2}},$$

用它可以得到由所有形如

$$p + q\sqrt{k}$$

的数,它们也形成一个域.称为 Q_1 的扩张,记为 Q_2,现在 p,q 可以是 Q_1 中的任意数,即 p,q 形如 $a + b\sqrt{2}$, a,b 为有理数.

从 Q_2 出发,我们还可以进一步扩充作图的范围.这种办法可以一直继续下去.用这种办法得到的数都是可用直尺圆规作出来的.

6.2.4 一般讨论

代数研究的对象是数,数偶(即坐标),一次方程式,二次方程式等.几何研究的对象是点,直线,圆,曲线等.通过坐标法,我们已将几何的对象与代数的对象紧密地联系在一起了.我们正在做的工作就是把几何对象化为代数对象.

现在面临一个这样的问题:用直尺圆规作出来的数是不是都在这个范围内呢?会不会超出这个范围呢?下面来回答这一问题.假定我们可用直尺圆规作出某个数域 F 中的所有数.

命题 2 只用直尺作不出数域 F 以外的数.

证明 设 $a_1, b_1, a_2, b_2 \in F$.过点 (a_1, b_1), (a_2, b_2) 的直线方程是

$$y - b_1 = \frac{b_2 - b_1}{a_2 - a_1}(x - a_1),$$

或
$$(b_1 - b_2)x + (a_2 - a_1)y + (a_1 b_2 - a_2 b_1) = 0.$$

它的系数是由 F 中的数作成的有理式.

今有两条以 F 中的数为系数的直线:

$$\alpha x + \beta y + \gamma = 0,$$
$$ax + by + c = 0,$$

解此联立方程，可得交点坐标：

$$x = \frac{b\gamma - c\beta}{b\alpha - a\beta}, y = \frac{c\alpha - a\gamma}{b\alpha - a\beta},$$

它们都是 F 中的数．这样一来，只用直尺作图不能使我们超出 F 的范围．

易见，用圆规可作出 F 以外的数．只需在 F 中取一 k，使 $\sqrt{k}$ 不在 F 中．我们能作出 $\sqrt{k}$，因而可作出所有形如

$$a + b\sqrt{k} \tag{4}$$

的数，其中 a,b 在 F 中．所有形如(4)的数形成一个域 F_1，它是 F 的扩域．

现在讨论反方向的问题．

命题 3 用圆规只能作出形如(4)的数．

证明 首先指出，圆规在作图中所起的作用只是确定一个圆与一条直线的交点或切点，或一个圆与另一个圆的交点或切点．通过解联立方程可以把交点或切点求出来．以 (ξ,η) 为中心，以 r 为半径的圆的方程是

$$(x - \xi)^2 + (y - \eta)^2 = r^2,$$

设 $\xi,\eta,r \in F$．将上式展开得，

$$x^2 + y^2 + 2\alpha x + 2\beta y + \gamma = 0;$$
$$\alpha = -\xi, \beta = -\eta, \gamma = \xi^2 + \eta^2 - r^2.$$

其中 α,β,γ 在 F 内．求圆与直线的交点或切点就是解联立方程组

$$x^2 + y^2 + 2\alpha x + 2\beta y + \gamma = 0,$$
$$ax + by + c = 0,$$

其中 $a,b,c \in F$．从第二个方程解出 y，代入第一个方程，得到一个二次方程

$$Ax^2 + Bx + C = 0,$$

其中 $A = a^2 + b^2, B = 2(ac - b^2\alpha - ab\beta), C = c^2 - 2bc\beta + b^2\gamma$. 其解为

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A},$$

它们可以化为形式 $p + q\sqrt{k}, p, q, k \in F$. 易见, $x \in F_1$. 交点 y 的坐标也有类似的公式. 这就是说, 圆和直线的交点的坐标都在扩域 F_1 中.

接着我们研究两个圆的交点或切点. 在代数上就是解二元二次联立方程:

$$x^2 + y^2 + 2\alpha x + 2\beta y + \gamma = 0,$$
$$x^2 + y^2 + 2\alpha' x + 2\beta' y + \gamma' = 0,$$

从第一个方程减去第二个方程, 得

$$2(\alpha - \alpha')x + 2(\beta - \beta')y + (\gamma - \gamma') = 0.$$

和前面一样, 把它与第一个圆的方程联立起来求出 x, y. 它们都不超出 F 的扩域 F_1.

无论是哪一种情形, 作图所产生的一个或两个新点的 x 坐标和 y 坐标其量的形式都是 $p + q\sqrt{k}, p, q, k \in F$. 在特殊情况下, $\sqrt{k}$ 本身也可以属于 F, 例如, 在有理数域中取 $k = 4$, 那么 $\sqrt{k} = 2$ 仍在有理数域中.

小结　1) 如果开始给定一些量, 那么从这些量出发, 只用直尺经有理运算可生成域 F 的所有量, 但不能超出域 F.

2) 用圆规能把可作图的量扩充到 F 的扩域 F_1 上. 构造扩域的过程可不断进行, 而得出扩域 $F_2, F_3, \cdots, F_n, \cdots$.

3) 可作图的量是而且仅仅是这一系列扩域中的数.

4) 可作图的数都是代数数(图 6 - 3).

6.2.5　代数知识

给定一个实数的子域 F, 设 F 是可构造域.

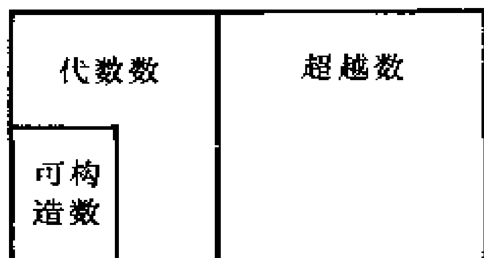


图 6 · 3

再设 $k > 0, k \in F$，但 $\sqrt{k} \notin F$. 我们记

$$F(k) = \{a + b\sqrt{k} : a, b \in F\}.$$

它是 F 的一个扩域. 例如，若取 $F = Q$，则

$$Q(2) = \{a + b\sqrt{2} : a, b \in F\},$$

$$Q(3) = \{a + b\sqrt{3} : a, b \in Q\}.$$

定义 由数 0 与 1 经过有限次加，减，乘，除（零不能作除数）以及正数的开方运算后得到的数称为二次不尽根.

例如，数 $1 + \sqrt{2 + \sqrt{5}}$ 就是一个二次不尽根.

定义 设 $F_0, F_1, \cdots, F_n$ 都是实数域的子域. 再设 $F_0 = Q$，且每一个域 F_k 都是域 F_{k-1} 的扩域（$k = 1, 2, \cdots, n$），则称 F_n 是 Q 的 n 阶的二次扩域.

例 构造数 $p = \sqrt{7} + \sqrt{3}$.

解 取 $F_0 = Q, k_0 = 7$，则 $\sqrt{7} \notin F_0$. 作

$$F_1 = \{a + b\sqrt{7} : a, b \in Q\}. \Rightarrow \sqrt{7} \in F_1.$$

取 $k_1 = 3$. $3 \in F_0 = Q \Rightarrow 3 \in F_1$. 但 $\sqrt{3} \notin F_1$. 作

$$F_2 = \{a + b\sqrt{3} : a, b \in F_1\}. \Rightarrow \sqrt{3} \in F_2.$$

取 $a = \sqrt{7} \in F_1 \Rightarrow \sqrt{7} \in F_2 \therefore \sqrt{3} + \sqrt{7} \in F_2$.

有了此例，我们在原则上就知道了如何构造一个扩域，使之包含我们所需要的元素.

不难看出，每个二次不尽根都属于某一阶的二次扩域.

设 $a,b,k\in F,k>0,\sqrt{k}\notin F$. 考虑数 $a+b\sqrt{k}\in F(k)$.

定义　设 $x=a+b\sqrt{k}$，则称数 $a-b\sqrt{k}$ 是 x 在 $F(k)$ 中的共轭数，记为 $\overline{x}=a-b\sqrt{k}$.

由定义立即看出，$x+\overline{x}=2a\in F$，即 $x\in F(k)\Rightarrow x+\overline{x}\in F$.

容易验证，共轭数具有如下的简单性质：

1) $x=0\Leftrightarrow\overline{x}=0$.　　2) $x\overline{x}=a^2-kb^2$.

3) $x=\overline{x}\Leftrightarrow x=a$.　　4) $z=x+y\Rightarrow\overline{z}=\overline{x}+\overline{y}$.

5) $z=xy\Rightarrow\overline{z}=\overline{x}\,\overline{y}$.　　6) $z=x^n\Rightarrow\overline{z}=(\overline{x})^n$.

这些性质与共轭复数的性质一样. 下一个定理也与共轭复数相同.

定理 1　设

$$P(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n$$

是一个 n 次多项式，它的系数属于某个域 F，F 由可构造数组成. 若 $x_0\in F(k)$ 是多项式 $P(x)$ 的一个根，则 $\overline{x_0}$ 也是 $P(x)$ 的一个根，即

$$P(x_0)=0\Rightarrow P(\overline{x_0})=0.$$

证　利用共轭的性质，在方程两边取共轭即可. 由

$$a_0x_0^n+a_1x_0^{n-1}+\cdots+a_{n-1}x_0+a_n=0,$$

可得

$$a_0\overline{x}_0^n+a_1\overline{x}_0^{n-1}+\cdots+a_{n-1}\overline{x}_0+a_n=0.$$

这就是要证的.

系　设 $a_1,a_2,a_3\in F$，F 是由可构造实数组成的域. 若三次方程

$$x^3+a_1x^2+a_2x+a_3=0\tag{5}$$

在扩域 $F(k)$ 中有一个根，则它一定在 F 中有一个根.

证　由代数基本定理，三次方程在复数域中有三个根，设它们是 x_1,x_2,x_3. 由韦达定理，

$$x_1+x_2+x_3=-a_1\in F.$$

若 $x_1\in F(k)$，则由定理 1，$\overline{x_1}$ 也是方程的根，不妨设 $x_2=\overline{x_1}$. 由共

轭的定义，$x_1+x_2\in F$．这样一来，$x_3=-a_1-(x_1+x_2)\in F$．定理得证．

注 若方程(5)在 F 中没有根，则它在 $F(k)$ 中也不会有根．

我们将把系应用到 F 是有理数域的情况，即假定方程(5)的系数都是有理数．在这种情况下，如果方程(5)在 $Q(k)$ 中有一个根，则它一定有一个有理根．所以该方程有有理根是方程有二次不尽根的必要条件．换言之，若该方程没有没有有理根，则它不会有二次不尽根．

不难看出，有理系数的 n 次代数方程可化为整系数的 n 次代数方程．事实上，只要乘上系数分母的最小公倍数就行了．我们还需要下面的定理．

定理 2 若整系数的 n 次方程

$$a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n=0 \tag{6}$$

有有理根 $\frac{a}{b}$（既约分数），则 a 是 a_n 的因数，b 是 a_0 的因数．

证 将 $\frac{a}{b}$ 代入方程(6)，得

$$a_0\frac{a^n}{b^n}+a_1\frac{a^{n-1}}{b^{n-1}}+\cdots+a_{n-1}\frac{a}{b}+a_n=0$$

$$a_0a^n+a_1a^{n-1}b+\cdots+a_{n-1}ab^{n-1}+a_nb^n=0$$

$$a(a_0a^{n-1}+a_1a^{n-2}b+\cdots+a_{n-1}b^{n-1})=-a_nb^n$$

由于 a 与 b 是互素的，所以 a 是 a_n 的因数．同样，用提出公因数 b 的方法可证明，b 是 a_0 的因数．

将上面结果应用到两个特殊方程上面去．

例 证明方程

$$x^3-2=0 \tag{7}$$

没有二次不尽根．

证 由定理 1 的系，只需证明方程没有有理根，因为方程没有

有理根，方程也自然不会有二次不尽根. 如果(7)有有理根$\frac{a}{b}$，则由定理2，a是2的因数，b是1的因数，因而，a只能取$\pm 1, \pm 2$，b只能取± 1. 这样$a/b = \pm 1, \pm 2$. 直接验证就知道它们都不是方程(7)的根. 这样一来，方程(7)没有有理根，从而也没有二次不尽根的解.

例 证明方程

$$8x^3 - 6x - 1 = 0 \tag{8}$$

没有二次不尽根的解.

证 如果方程(8)有有理根$\frac{a}{b}$，则a是1的因子，b是8的因子. 这样一来，方程(8)的有理根不外是$\pm 1, \pm \frac{1}{2}, \pm \frac{1}{4}, \pm \frac{1}{8}$ 直接验证知道它们都不是. 由定理1的系，方程(8)没有二次不尽根的解.

6.2.6 三大难题的解

有了上面的准备，我们来解三大几何难题.

1) 立方倍积问题 设给定的立方体是单位立方体，它的边长是单位长度. 若体积为这立方体体积两倍的立方体的边长是x，则

$$x^3 - 2 = 0.$$

如果立方倍积问题可解，则我们一定能用直尺和圆规构造出长度为$\sqrt[3]{2}$的线段. 但是前面已经证明方程(7)没有有理根，也没有二次不尽根. 这样一来，立方倍积问题是不可解的.

2) 三等分任意角 我们现在要证明只用直尺和圆规三等分任意角一般说来是不可能的. 当然，像90°和180°那样的角是可以三等分的. 我们要说明的是，对每一个角的三等分都有效的办法是不存在的. 为了证明这一点，只要证明有一个角不能三等分就足够了，因为一个合理的一般方法必须适用于每一种情况. 因此如果我们能够证明60°角只用直尺和圆规不能三等分，那就证明了一般方法是不存在的.

如图6－4所示，我们从60°角着手. 设$\angle QOP = 60°$，并设线段

OP 的长度为 1. 假定三等分任意角是可能的. 如图设 $\angle ROP = \theta = 20^\circ$. 那么,点 R 的纵坐标一定是有理数或二次不尽根. 这相当于说 $\cos\theta = 1/OR$ 是有理数或二次不尽根. 我们需要公式(见第 2 章 §2.3 公式(10))

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta.$$

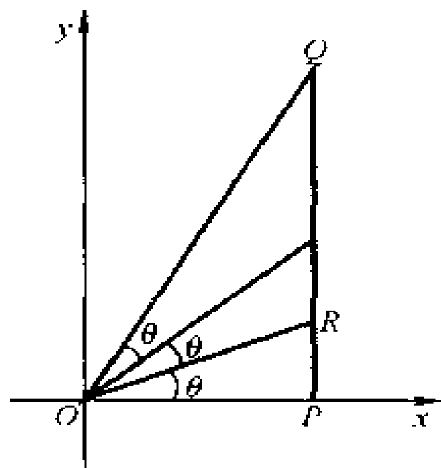


图 6 - 4

现在 $\cos 3\theta = \cos 60^\circ = \frac{1}{2}$,所以

$$4\cos^3\theta - 3\cos\theta = \frac{1}{2}.$$

令 $x = \cos\theta$ 并代入上式,得到

$$8x^3 - 6x - 1 = 0.$$

这正是前面讨论过的方程(8). 这个方程没有有理根,也没有二次不尽根. 这说明我们的假定是不对的. 这就证明了三等分任意角是不可能的.

我们知道,60° 角可作,因而正六边形可作. 若 60° 角可三等分,则正 18 边形可作,从而正 9 边形也可作. 刚才已经证明,60° 角不可三等分,因而正 9 边形不能只用直尺和圆规作出来.

3) 化圆为方　考虑半径为 1 的单位圆,它的面积为 π,现在构造一个边长为 x 的正方形,它的面积为 π,于是 $x^2 = \pi$, $x = \sqrt{\pi}$. 由于 $\sqrt{\pi}$ 是一个超越数,所以它不可能是二次不尽根,因此"化圆为方"的问题是不可解的.

自然对数的底 e 与 π 都是超越数. 证明它们是超越数是困难的,吸引着许多数学家付出巨大的劳动去进行研究. 直到 1873 年埃尔米特才给出了 e 是超越数的证明. 他认为证明 π 的超越性更困难,而不敢去尝试. 他给友人的信中写道:"我不敢去试着证明 π 的超越性. 如果其他人承担这项工作,对于他们的成功没有比我更高兴的人了,但

是请相信我,我亲爱的朋友,这决不会不使他们花去一些力气". 九年之后,林德曼在 1882 年用实质上与埃尔米特相同的方法证明了 π 的超越性.

用直尺圆规不能三等分任意角是指,直尺只能用来画直线,圆规只能用来画圆,而不能作别的用途. 只有在这时定理才是正确的. 如果容许直尺作别的用途,那么可作图的范围就可大大扩大. 为了说明这一点,下面举一个三等分任意角的例子. 这个例子是在阿基米德的著作中发现的.

问题　设 $\angle\alpha$ 是一个任意角. 求 $\angle\beta = \frac{1}{3}\angle\alpha$.

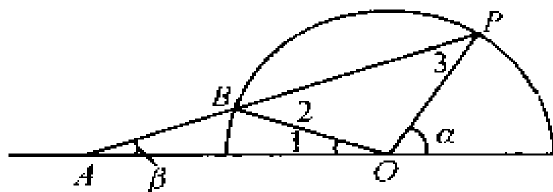


图 6－5

解　在图 6－5 中,向左边延长 $\angle\alpha$ 的底边. 以 O 为心,以 r 为半径作半圆. 在直尺上标出 A,B 两点使 $AB = r$. 让 B 点保持在半圆周上,直尺通过点 P,滑动并转动直尺,让 A 点落在底边延长线上. 直尺和 $\angle A$ 的底边形成一个角:$\angle\beta$. 这就是所求的角.

证　由图 6－5,$AB = BO \Rightarrow \angle 1 = \angle\beta$. 又

$$\angle 2 = \angle 3 = 2\angle\beta.$$

由外角定理,

$$\angle\alpha = \angle\beta + \angle 3 = 3\angle\beta.$$

这就三等分了任意角.

所以容许直尺作别的用途,三等分任意角就是可能了.

习 题

1. Q_1 是所有形如(3)的数构成的域,取 $k=\sqrt{2}$,证明其扩域中的数具有形式 $p+q\sqrt[4]{2}$,其中 $p,q\in Q_1$.

2. 直线

$$x+\sqrt{2}y-1=0,2x-y+\sqrt{2}=0$$

的系数在域(3)中.计算它们的交点坐标,并验证它的形式为(3).

§6.3 正多边形作图

在讲费马数的时候曾讨论过正多边形作图,现在对正多边形作图再作些补充讨论.首先回顾高斯定理.

高斯定理 对奇数 n,当且仅当 n 是一个费马素数,或者若干个费马素数的乘积时,正 n 边形才能用直尺和圆规作图.

$n=2$ 的情况是平凡的,不包含在定理之中.我们把基本结果罗列如下.

1) 用直尺和圆规可等分一个任意角.通过平分 180° 角,可作出正四边形.进而作出正 2^n 边形,$n=2,3,\cdots$.

2) $n=3$ 是第一个费马素数:$2^{2^0}+1=3$.我们会作正三角形,从而会作正 $3\cdot 2^n$ 边形.

3) $n=5$ 是第二个费马素数:$2^{2^1}+1=5$.我们来回顾正五边形的作图.

我们从正十边形开始.如图 6－6 所示,假定一个正十边形内接于半径为1的圆.设正十边形的边长为 x.在图上 $AB=x$,它所对应的中心角 $\angle AOB=36°$.从而 $\angle OAB=\angle OBA=72°$.作 $\angle OAB$ 的平分线 AC,交 OB 于 C.这时 $\triangle ACB$,$\triangle AOC$ 都是等腰三角形,并且

$\triangle AOB \sim \triangle ACB$. 由此可得，

$$\frac{1}{x} = \frac{x}{1-x} \Rightarrow x^2 + x - 1 = 0.$$

解方程得 $x = (\sqrt{5} - 1)/2$（另一根是负的，舍去它）.

由前面的讨论知，$\sqrt{5}$ 可作，$(\sqrt{5} - 1)/2$ 也可作. 因而我们可以作出 x，由此可作出正十边形，隔点相连就可作出正五边形. $\sqrt{5}$ 这样作：作一直角三角形，使直角边分别是 1 和 2，则斜边就是 $\sqrt{5}$.

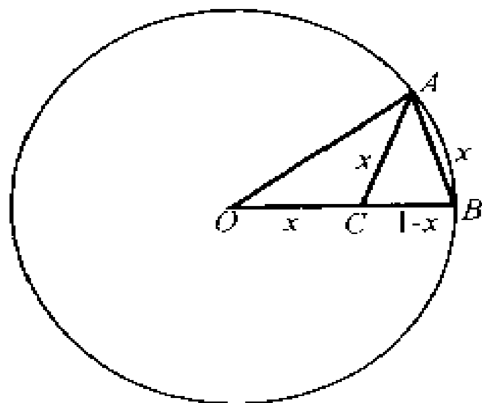


图 6 6

4）前面曾提到，只用直尺和圆规不可能作出正七边形. 现在我们给出证明. 正七边形的顶点是方程

$$z^7 - 1 = 0 \tag{1}$$

的根. 由

$$z^7 - 1 = (z - 1)(z^6 + z^5 + \cdots + z + 1) = 0 \tag{2}$$

可知 $z = 1$ 是方程的一个根，其它根满足方程

$$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0. \tag{3}$$

用 z^3 除(3)，得

$$z^3 + \frac{1}{z^3} + z^2 + \frac{1}{z^2} + z + \frac{1}{z} + 1 = 0,$$

通过简单的代数运算，可以把它化为

$$\left(z + \frac{1}{z}\right)^3 + \left(z + \frac{1}{z}\right)^2 - 2\left(z + \frac{1}{z}\right) - 1 = 0.$$

令 $y = z + 1/z$，则上式化为

$$y^3 + y^2 - 2y - 1 = 0. \tag{4}$$

由于 z 是 1 的 7 次根，故可设

$$z = \cos\varphi + \mathrm{i}\sin\varphi, \varphi = \frac{2\pi}{7}.$$

而
$$\frac{1}{z} = \cos\varphi - i\sin\varphi.$$

两式相加就得出 $y = 2\cos\varphi$

如果我们能作出 y，就能作出 $\cos\varphi$，反之亦然. 如果我们证明了 y 是不能用尺规作图的，我们也就证明了正七边形是不能用尺规作图的. 根据 §6.2 定理 1 的系，只需证明方程(4) 没有有理根.

假定方程(4) 有一个有理根 r/s，其中 r 与 s 没有公因数. 则我们有

$$\left(\frac{r}{s}\right)^3 + \left(\frac{r}{s}\right)^2 - 2\frac{r}{s} - 1 = 0,$$

或
$$r^3 + r^2s - 2rs^2 - s^2 = 0,$$

即
$$r^3 + r^2s - 2rs = s^2.$$

由此可看出，r^3 中含有因子 s，而 s^3 中含有因子 r. 由于 r，s 没有公因数，所以它们必须等于 ±1. 这样一来，如果 y 是有理数的话，它只能取 1 或 −1. 把这些数代入方程，可知它们都不是根. 这就证明了正七边形是不能用尺规作图的.

5) 正 9 边形是不能用尺规作图的，因为 $9 = 3\times 3$，是两个相等的费马素数相乘. 正 11 和 13 边形也是不能用尺规作图的，因为这两个数不是费马素数.

6) 正 17 边形可用尺规作图的. 这是高斯的一大贡献. 他把从古希腊开始 2000 年来没有进展的问题大大推进了一步. 高斯证明了可以用直尺圆规作正 17 边形，并且自己完成了这个作图. 对这个作图有兴趣的读者可参考作者的《复数，复函数及其应用》(湖南教育出版社) 一书.

§6.4　平行公设引起的思考

欧几里得公设中的 1)—4) 都很容易地被人们接受了. 唯独公

设5)从一开始就受到人们的怀疑.实际上受到质疑的不是欧几里得的陈述,而是将它列为公设.古代就有人说:"它完全应该从公设中剔除,因为它是一条定理…".的确,这一公设看起来确像一条命题,它的陈述性语言就占了一大半.

对第五公设的研究导致了非欧几何的诞生,这段历史大致可以分为四个时期:1)寻求第五公设的证明时期;2)非欧几何的孕育时期;3)非欧几何的诞生时期;4)非欧几何的确认时期.

6.4.1 从《几何原本》诞生到18世纪

这期间有两种途径:一是用更为自明的命题代替第五公设;二是企图从欧几里得的其它几个公设中推导出第五公设来.多年来提出的替代公设的有:

1)存在一对同平面的直线彼此处处等距离.

2)过已知直线外的已知点只能作一条直线平行于已知直线.

3)存在一对相似但不全等的三角形.

4)如果有一个四边形有一对对边相等,并且它们与第三边构成的角均为直角,则余下的两个角也是直角.

5)如果四边形有三个角是直角,则第四个角也是直角.

6)至少存在一个三角形,其三角和等于二直角.

7)过任何三个不在同一直线上的点可作一圆.

8)三角形的面积无上限.

今天中学几何课本中最喜欢用的是上述的2).人们把它归功于苏格兰物理学家和数学家普雷菲尔(J.Playfair 1748—1819).

多少个世纪以来,从欧几里得的其它假定推出第五公设的尝试是如此之多,差不多够一个军团,所有这些尝试均告失败,其中绝大多数或迟或早依靠了与该公设本身等价的隐含的假定.这些工作的绝大多数对数学思想的发展没有什么现实意义.直到1733年意大利人萨谢利(G.Saccheri)才做了关于第五公设(现在也叫平行公设)值得注意的研究成果.

6.4.2　非欧几何的孕育时期

在第一阶段虽然没有得出实质性的成果，但却积累了大量的经验与资料.第二阶段的主要代表人物是萨谢利，兰伯特和勒让德等人.

萨谢利1667年出生于意大利的圣拉蒙.少年早熟，23岁就完成了其耶稣会神职的见习期，然后一直在大学担任教学职务.在米兰的耶稣会学院中讲授修辞学、哲学和神学时，他读了欧几里得的《原本》，并且醉心于强有力的归谬法.稍迟，在都灵教哲学时，他发表了《逻辑证明》(Logica demonstrativa)一书，其中的主要改进是应用归谬法来处理形式逻辑.几年以后，在帕维亚大学任数学教授时，他把他喜爱的归谬法用于对欧几里得平行公设的研究.他为这项工作做了很好的准备：在其较早的关于逻辑的著作中已经灵活地处理了定义和公设这类事物.他还熟悉其他人讨论平行公设的著作，并且成功地指出了在纳瑟·埃得和沃利斯的尝试中的谬误.

在曾经研究过否定欧几里得平行公设会得到什么样的结果的人当中，萨谢利显然是第一个试图应用归谬法来证明这一著名公设的.他的研究结果写在题为《排除任何谬误的欧几里得》的一本小书中，这本书是于1733年作者去世前几个月在米兰出版的.在这一著作中，萨谢利承认《原本》的前二十八命题；证明这些命题不需要第五公设.借助于这些定理，他研究等腰双直角四边形，即四边形 $ABCD$(见图6－7)，在其中，$AC=BD$，且$\angle A$和$\angle B$均为直角.作对角线AD和BC，再利用全等定理(包含在欧几里得的前二十八个命题中)，萨谢利容易地证明了：$\angle C=\angle D$.但无法确定这两个角的大小.当然，作为第五公设的推论，可推出这两个角均为直角，但是他不想采用此公设的假定.因而这两角可能均为直角，或均为钝角，或均为锐角.萨谢利在这里坚持了开放的思想，并且把这三种可能性命名为直角假定、钝角假定和锐角假定.他的计划是，以证明后两个假定导致矛盾来排除这两种可能，然后根据归谬法就只剩下第一个假定了.但是这个假定等价于欧几里得第五公设.这么一来，平行公设

就被证明了,欧几里得假定的缺陷就被排除了.

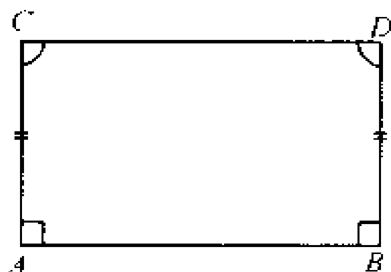


图 6 – 7

萨谢利以其娴熟的几何技巧和卓越的逻辑洞察力证明了许多定理.现将其中较重要者列举如下:

1) 直角假定 ⇔ 三角形内角和等于两直角.
⇒ 给定一条直线和线外一点,过该点有一条直线与该直线不交.
⇒ 立于固定直线上的定长垂线的顶点轨迹是一条直线.

2) 钝角假定 ⇔ 三角形内角和大于两直角
⇒ 没有平行线
⇒ 立于固定直线上的定长垂线的顶点轨迹是凸曲线.

3) 锐角假定⇔ 三角形内角和小于两直角.
⇒ 给定一条直线和线外一点,过该点有无穷多条直线与该直线不相交.
⇒ 立于固定直线上的定长垂线的顶点轨迹是凹曲线.

可惜的是,萨谢利的著作没有带来多大影响,并且没过多久就被人们遗忘了.直到 1889 年才被他的同胞贝尔特拉米(E. Beltrami, 1835—1900)戏剧般地给与了新的生命.

1766年，萨谢利发表其著作之后33年，瑞士的H·兰伯特(Lambext,1728－1777)写了一本标题为《平行线理论》的著作，作了类似的研究；不过，这部著作在兰伯特死后11年才发表.兰伯特选作基础图形的是三直角四边形，它可以看作是：一连接萨谢利等腰双直角四边形两底中点而形成的"半个萨谢利四边形".和萨谢利一样，兰伯特按照三直角四边形的第四个角是直角、钝角或锐角作了三个不同的假定.

兰伯特和萨谢利都在钝角和锐角的假定下推演出了不少命题，然而，兰伯特走得更远.例如，和萨谢利一样，他证明了：在这三个假定下分别可推出三角形内角和等于、大于或小于两个直角；然而，他进一步证明了：在钝角假定下大于两个直角的超出量和在锐角假定下小于两个直角的亏量均与三角形的面积成正比.他看到由钝角假定推出的几何与球面几何的类似之点：在球面几何中，三角形的面积与其球面角盈成正比.他还猜测：由锐角假定推出的几何也许能在虚半径的球上被证实；这也猜对了.

兰伯特和萨谢利一样，以默认直线为无限长这个假定来取消钝角假定.

兰伯特的几何观点是十分先进的.他认识到任何一组假设如果不导致矛盾的话，一定提供一种可能的几何.

用归谬法证明欧几里得平行公设的第三个卓越的贡献是由法国著名数学家勒让德(A.M.Legendre,1752—1833)作出的.他对一特殊的三角形的内角和做出三个不同的假定：等于、大于或小于两直角.他隐含地承认直线的无限性，因而取消第二假定；但是，尽管他作了种种尝试，还是没法排除第三个假定.

勒让德的另一个重要贡献是他的备受欢迎的著作《几何学基本原理》(Elements de geometrie)，于1794年出第一版.他对欧几里得《原本》作了教学法上的改进，重新安排和简化了许多命题，成为现在流行的形式.

施韦卡特(F.K.Schweikart, 1780—1859),一位法学教授,业余研究数学,他更迈进了一步,研究非欧几里得几何.1816年,他写了一份备忘录,于1818年送交高斯征求意见.他区分了两种几何:欧几里得与假设三角形内角之和不是两直角的几何.他称后一种几何为星空几何,因它可能在星空内成立.它的定理都是萨谢利和兰伯特根据锐角假设建立的定理.

陶里努斯(F.A.Taurinus,1794—1874),施韦卡特的外甥,接受舅父的建议继续研究星空几何.虽然他证实了一些新结果,但他仍做出结论:只有欧几里得几何对物质空间是正确的,而星空几何只是逻辑上相容.

兰伯特、施韦卡特、陶里努斯这三人,还有当时一些其他人都承认欧几里得平行公设不能证明.这三人也都注意到实球面上的几何具有钝角假设为基础的性质,而虚球面上的几何则具有以锐角假设为基础的性质.这样,所有三个人都认识到非欧几里得几何的存在性,但他们都失去一个基本点,即欧几里得几何不是唯一的在经验能够证实的范围内来描述物质空间的性质的几何.

这段历史已清楚地表明非欧几何已是躁动于母腹中的婴儿了.

6.4.3 非欧几里得几何的诞生

从前面我们看到,尽管经过长时间的艰苦努力,萨谢利、兰伯特和勒让德还是没有能以锐角假定为前提推出矛盾.在此假定下找不到矛盾,没有什么可惊讶的,因为现在我们已经知道,由某一组基本假定加上锐角假定推出的那套几何,和由同样的一组假定加上直角假定推出的欧几里得几何一样,是自相容的.换言之,平行公理不能作为定理从欧几里得的其它假定推出,它独立于其它那些假定.对于两千年来受传统偏见的约束,坚信欧几里得几何无疑是唯一可靠的几何,而任何与之矛盾的几何系统绝对是不可能相容的人来说,承认这样一种可能是要有不寻常的想象力的.

高斯是真正预见到非欧几何的第一人.不幸的是,毕其一生高斯

没有关于此命题发表什么意见.他的先进思想表示在他通过与好友的通信、对别人著作的几份评论,以及在他死后从稿纸中发现的几份札记中.虽然他克制住自己,没有发表自己的发现,但是他竭力鼓励别人坚持这方面的研究.把这种几何称为非欧几何的正是他.

预见到非欧几何的第二人是J·鲍耶(J. Bolyai),匈牙利人.他是数学家F·鲍耶的儿子.F·鲍耶与高斯有长期的亲密的友谊.小鲍耶的这项研究受到他父亲的很大启发,因为老鲍耶早就对平行公设问题感兴趣.早在1823年J·鲍耶就开始理解摆在他面前的问题的实质.那年他给父亲写了一封信,说明他热衷于这项工作,并强调说:"我要白手起家创造一个奇怪的新世界."J·鲍耶称他的非欧几何为绝对几何,他写了一篇26页的论文《绝对空间的几何》,出版时作为附录附于他父亲的《为好学青年的数学原理论著》,出版于1823—1833年间.J·鲍耶似乎在1825年已建立起非欧几何的思想.

虽然人们承认高斯和J·鲍耶是最先料想到非欧几何的人,但是俄国数学家罗巴切夫斯基(N. I. Lobatchevsky,1793—1856)实际上是发表此课题的有系统的著作的第一人.罗巴切夫斯基一生中的大部分时间是在喀山度过的,先是学生,后来任数学教授,最后是当校长.他关于非欧几何的最早论文就是于1829—1830年在《喀山通讯》上发表的,比鲍耶著作的发表早二到三年.这篇论文在俄国没有引起多大注意,因为是用俄文写的,实际上在别处也没有引起多大注意.

他发展的非欧几何现今被称为罗巴切夫斯基几何.他赢得了"几何学上的哥白尼"的称号.

在罗巴切夫斯基和鲍耶的著作发表若干年后,整个数学界才对非欧几何这个课题给予更多的注意.几十年后这项发现的真正内涵才被理解.下一个重要任务是证明新几何的内在相容性.

6.4.4 罗巴切夫斯基的解答

罗巴切夫斯基在他的著作《新几何原本》里用以下这段话描述了他所给出的第五公设的解答要点:

“大家知道，直至今天为止，几何学中的平行线理论还是不完全的.从欧几里得时代以来，两千年来的徒劳无益的努力，促使我怀疑在概念本身之中并未包括那样的真实情况，它是大家想要证明的，也是可以像别的物理规律一样单用实验(譬如天文观测)来检验的.最后，我肯定了我的推测的真实性，而且认为困难的问题完全解决了.我在1826年写出了关于这个问题的论证.”

这段话集中了罗巴切夫斯基的新观点，不仅给出了关于第五公设的问题的解答，而且使几何学的全部注意力转移到新的方面，甚至还不单是几何学如此.他的解答实质上包含这样三个方面：

1) 公设是不能证明的；

2) 几何学的其它基础命题添上否定公理以后，可以展开一种与欧几里得几何不同的、逻辑上完整而富有内容的几何学；

3) 这种或那种逻辑上可能的几何学的结论，在应用到现实空间时的正确性只有用实验来作检验.逻辑上可能的几何学不应该当作任意的逻辑体系来研究，而应该作为促成发展物理理论的可能途径和方法的理论来研究.

这后一观点在后来的爱因斯坦的相对论中得到证实.

6.4.5　非欧几何的相容性

虽然罗巴切夫斯基和鲍耶在他们对于以锐角假定为基础的非欧几何的广泛研究中没有遇到矛盾；虽然他们甚至相信，不会产生矛盾；但是仍然有这种可能，如果这类研究充分地继续下去，会出现矛盾，或不相容.平行公设对于欧几里得几何其它公设的独立性，无疑，要在锐角假定相容性做出之后才能成立.这些没有多久就做到了.那是贝尔特拉米，凯利，F·克莱因，庞加莱等人的工作.办法是在欧几里得几何内建立一个新几何的模型，使得锐角假定的抽象发展在欧几里得空间的一部分上得到表示.于是，非欧几何中的任何不相容性会反映此表示的欧几里得几何中的对应的不相容性.这种证明是相对相容性的一种；如果欧几里得几何是相容的，则可证明罗巴切夫斯

基几何是相容的.当然,每个人都相信欧几里得几何是相容的.

罗巴切夫斯基的非欧几何的相容性的成果之一是,古老的平行公设问题的最终解决.相容性确定了下述事实:平行公设独立于欧几里得几何的其他假定,把此公设当定理,由其它假定推出它的可能性是不存在的.因为如果平行公设可被推出,则这个与罗巴切夫斯基平行公设矛盾的结果会在非欧几何体系中构成不相容.非欧几何的相容性还有一些后果,其影响远远超过了平行公设问题的解决.其中的一个重要后果是,几何学从传统的模型中解放了出来.几何学的公设,对数学家来说,仅仅是假定,其物理上的真与假用不着考虑;数学家可以随心所欲地选取公设,只要它们彼此相容.当数学家采用公设这个词时,并不包含"自明"或"真理"的意思.有了发明"人造的几何"的可能.

事实上,罗巴切夫斯基几何的相容性不仅解放了几何学,对于整个数学也有类似的影响,数学显现为人类思想的自由创造物.

6.4.6 黎曼的非欧几何

我们已经看到,钝角假定被所有在此课题上探索过的人所抛弃,因为它与直线无限长的假定相矛盾.认出以钝角假定为基础的第二种非欧几何的是德国数学家 G.F.B.黎曼(Riemann,1826—1866).这是他在1854年讨论无界和无限概念时得到的成果.虽然欧几里得的公设2断言:直线可被无限延长,但是,并不必定蕴涵直线就长短而言是无限的,只不过是说:它是无端的或无界的.例如,连接球上两点的大圆的弧可被沿着该大圆无限延长,使得延长了的弧无端,但确实就长短而言它不是无限的.现在我们可以设想:一条直线可以类似地运转,并且,在有限的延长之后,它又回到它本身.由于黎曼把无界和无限的概念分辨清了,可以证明,人们能实现满足钝角假定的一种内相容的几何,如果欧几里得的公设1,2和5作如下修正的话:

1) 两个不同的点至少确定一条直线.

2) 直线是无界的.

3) 平面上任何两条直线都相交.

这第二种非欧几何通常被称作黎曼非欧几何.

仔细地阅读欧几里得的公设 1 和公设 2,就会知道,它们实际上说的正是公设 1) 和 2) 所说的意思.

由于罗巴切夫斯基的和黎曼的非欧几何的发现,几何学从其传统的束缚中解放出来了,从而为大批新的、有趣的几何的发现开辟了广阔的道路.这些新几何有:非阿基米德几何,非笛沙格几何,黎曼几何,非黎曼几何,有限几何(它只包含有限多的点、线和面),等等.这些新几何并不是毫无用处的.例如在爱因斯坦发现的广义相对论的研究中,必须用一种非欧几何来描述这样的物理空间,这种非欧几何是黎曼几何的一种.再如,由 1947 年对视空间(从正常的有双目视觉的人心理上观察到的空间)所做的研究得出结论:这样的空间最好用罗巴切夫斯基非欧几何来描述.

6.4.7 欧氏几何与非欧几何

由于平行公设的不同而带来了欧氏几何与非欧几何的一些本质不同,例如,在罗巴切夫斯基的几何中三角形的内角和总小于 180°,就是一个著名的例子.这里再列举一些本质的不同:

半径无限大的圆周的极限不是直线,而是一种曲线,叫作极限圆;通过不在一条非欧直线上的三点,并不总能作一个非欧圆,而能做的或者是非欧圆,或者是极限圆,或者是等距线(即与一条非欧直线等距离的点组成的线);

不存在面积任意大的非欧三角形;

两个非欧三角形相似就全同;

毕达哥拉斯定理不成立;等等.

这里需要特别指出的是,在充分小的区域内非欧几何与欧氏几何的差异是非常小的;区域越小,这种差异就越小,例如在充分小的三角形里,普通三角学的公式是相当精确地描述了三角形内边与角的关系.在我们现实生活的领域内,我们还无法确定,究竟是欧氏几

何,还是非欧几何更符合我们的现实空间.

罗巴切夫斯基的几何也称为双曲几何,高斯与鲍耶的几何也属于这一几何,即三角形内角小于 180° 的几何.为了理解这种几何的某些古怪论点,我们不妨看一看这种几何对某些命题的证明.

定理(角角角)　若一个三角形分别与另一个三角形的三个角相等,则这两个三角形全等.

注　在欧氏几何中两个三角形的三个角分别相等,则这两个三角形相似,但它们不一定全等.两个不同的相似三角形可能有不同的边长,不同的面积,但是在非欧几何里却出现了奇怪的现象:两个三角形相似就全等.

欧几里得的全等三角形的定理出现在第五公设之前.也就是说,它们没有用到第五公设.因而这些定理在非欧几何中依然有效.现在我们就用在欧氏几何与非欧几何中都成立的全等定理,来证明上面的奇怪定理.

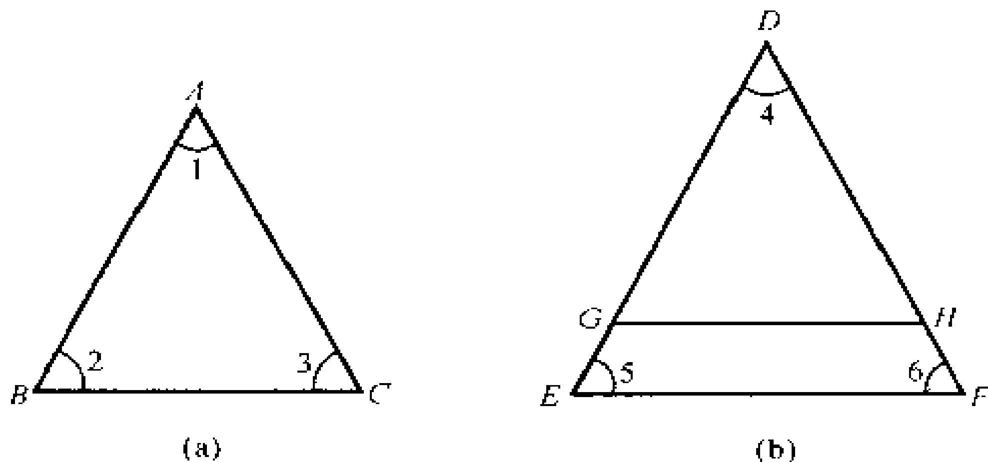


图 6　8

证　设 $\triangle ABC$,$\triangle DEF$(见图 6－8)满足条件 $\angle 1=\angle 4$,$\angle 2=\angle 5$,$\angle 3=\angle 6$,要证 $\triangle ABC\cong\triangle DEF$.根据角边角定理,我们只

需证 $AB = DE$. 今用反证法. 设:$AB < DE$. 作 $DG = AB$, 作 $\angle DGH = \angle 2$. 由角边角定理, $\triangle ABC \cong \triangle DGH$. 从而

$$\angle DGH = \angle 2 = \angle 5, \angle DHG = \angle 3 = \angle 6.$$

今研究四边形 $GEFH$. 在四边形 $GEFH$ 中,

$$\angle EGH = 180^\circ - \angle DGH = 180^\circ - \angle 5,$$

$$\angle FHG = 180^\circ - \angle DHG = 180^\circ - \angle 6.$$

在四边形 $GEFH$ 中, 四内角之和为

$$(180^\circ - \angle 5) + 180^\circ - \angle 6 + \angle 5 + \angle 6 = 360^\circ. \quad (1)$$

连接 GF 将四边形分为两个三角形. 每个三角形的内角和都小于 180°, 所以两个三角形内角和小于 360°. 这与与(1)矛盾. 这说明, 假定 $AB \neq DE$ 不正确, 所以 $AB = DE$. 再用角边角定理就知道,

$$\triangle ABC \cong \triangle DEF.$$

证明中用的关键事实是, 在非欧几何中每个三角形的内角和都小于 180°. 由此我们还可以看出, 在非欧几何中, 不同的三角形其内角和也不相同. 在欧氏几何中, 已知三角形的两个角可确定它的第三个角. 在非欧氏几何中, 已知三角形的两个角不能确定它的第三个角. 当时鲍耶说, 他创造了一个"全新的世界"现在不难体会, 他的确创造了一个全新的世界.

6.4.8 爱尔兰根纲领

正如前面指出的, 由于非欧几何的诞生, 几何学从其传统的束缚中解放出来了, 从而大批新的几何学诞生了. 于是出现了这样的问题: 什么是几何学? 几何学是研究什么的?

1872 年, 在爱尔兰根大学哲学教授评议会上, F. 克莱因(1849—1925)按照惯例作其专业领域的就职演讲. 演讲以他本人和挪威数学家 S. 李(1842—1899)在群论方面的工作为基础, 给"几何学"下了一个著名的定义. 就其本质而言, 是对当时存在的几何学进行了整理, 并为几何学的研究开辟了新的, 富有成果的途径. 这个演讲连同他提倡的几何学研究的规划, 已成为人们所熟悉的爱尔兰根

纲领.

克莱因的基本观点是,每一种几何都由变换群所刻画,并且每种几何要做的就是考虑这个变换群下的不变量.此外,一种几何的子几何就是考虑原来变换群的子群下的一族不变量.在这个定义下,相应于给定变换群的几何的所有定理仍然是子群中的定理.

克莱因也提出对一一对应连续变换下具有连续逆变换的不变量进行研究.这是现在叫做同胚的一类变换,在这类变换下不变量的研究是拓扑学的主题.把拓扑学作为一门重要的几何学科,这在1872年是一个大胆的行动.

克莱因的综合与整理指引了几何思想有50年之久.

按照克莱因的说法,存在七种相关的平面几何,其中包括欧几里得几何,双曲几何和椭圆几何.1910年英国数学家沙默维尔(D.M.Y.Sammerville)作了进一步细分,把平面几何的数目从七种增加到九种.

但是,不是所有的几何都能纳入到克莱因的分类方案中的.虽然克莱因的观点不能无所不包,但它确能给大部分的几何提供一个系统的分类方法,并提示很多可供研究的问题.

他所强调的变换下不变的观点已经超出数学之外而进入到力学和理论物理中去了.变换下不变的物理问题,或者物理定律的表达方式不依赖于坐标系的问题,在人们注意到麦克斯韦尔方程在洛仑兹变换(仿射几何的四维子群)下的不变性后,在物理思想中都变得很重要.这种思想路线引向了相对论.

回顾几何与代数的差别,我们可以这样说,几何学基本上是研究不变量的,而代数学基本上是研究结构的.

6.4.9 各种几何与物理空间

欧氏几何诞生最早,已有二千多年的历史.绝大多数人从小接触的就是欧氏几何.这不但对普通的平民百姓而言是这样,就是在数学家中,即使在20世纪的今天,懂得非欧几何的人也是少数.一个根深蒂固的思想是,我们生活的空间是欧氏空间.或者说,欧氏几何是物

理空间的几何，是关于空间的真理. 这种思想在一百多年前，非欧几何刚刚诞生的时期更是占绝对的统治地位. 以至于在许多年中，与之相悖的任何思想，包括高斯的，都被拒之门外. 数学家康托尔曾这样评述这种无知的保守：一旦错误的结论被广泛接受，那么它将不会轻易地放弃，而且对它懂得越少，则它的地位越牢固. 罗巴切夫斯基和鲍耶的著作发表后三十年左右的时间中，除了少数几个数学家外，几乎所有数学家都对其置之不理，它们被视为异端邪说. 有些数学家并不否认它们的逻辑上的一致性，另一些则相信它们必定包含着矛盾因而毫无价值. 几乎所有的数学家都坚持相信，物理空间的几何必须是欧氏几何. 1855 年高斯死后（此时他的声望已无人可比），他的笔记中的材料被公之于众. 黎曼于 1854 年写就的论文在 1868 年发表时使得许多数学家相信非欧几何也可以是物理空间的几何，单是还有别的几何存在就已是一个令人震惊的事实了，然而更令人震惊的是，你不再知道哪个是正确的，或者究竟有没有正确的.

非欧几何及其隐含的关于几何真理性的内容逐渐被数学家们所接受. 但并不是由于它的适用性的任何论据被加强了，而是正如普朗克，这位量子力学的奠基人在本世纪初所说的："一个新的科学真理并不是靠说服它的对手并使其看见真理之光取胜，而是由于它的对手死了，新的一代熟悉它的人成长起来了."

所有这些奇怪的几何都可与欧氏几何匹敌，甚至有可能取而代之. 这种想法乍听起来很是荒谬，但是高斯接受了这一可能性. 无论他是否确实使用了他在 1827 年写的论文中记录的测量方法来检验非欧几何的适用性，他是第一个肯定非欧几何的真理性的人.

根据他的一篇传记可知，高斯曾经试图检验这一观点. 他注意到在欧氏几何中三角形内角和为 180°，而在非欧几何中三角形内角和小于 180°，他曾花了几年时间对汉诺威王国进行测量，并记录了数据. 因此有可能他用这些数据来测量三角形的内角和. 在 1827 年写的一篇著名的论文中，高斯注意到由布诺肯山（Brocken）、霍赫海根

山(Hohehagen)和英色伯格山(Inselberg)三座山峰构成的三角形内角和为 180°15″. 这什么也证明不了,因为测量误差远大于 15″,也许正确的和不会超过 180°,高斯一定意识到这个三角形太小了. 因为在他的非欧几何中,三角形内角和与 180° 的偏离程度正比于它的面积. 只有非常巨大的三角形,比如在天文学研究中的三角形,才能显示出明显的偏离. 然而高斯还是相信这门新的几何和欧氏几何一样有实用性.

罗巴切夫斯基也考虑了他的几何在物理空间中的应用,而且确实给出了证据,说明它可用于非常大的几何图形. 因此,到了 19 世纪 30 年代,非欧几何已不仅仅是少数几个人接受了,而且它在物理空间的适用性被认为至少是可能的了.

第七章　同余理论及其应用

数学作为一个创造性的学科，按三个基本步骤运行：1) 体验一个问题，并从中发现一个模式；2) 定义一个符号系统来表达这一模式；3) 把这个符号系统组织为一个系统的语言.

G.C.M. Report

在数学中，要紧的不是记号而是概念.

C.F. 高斯

同余理论是初等数论的一个重要的组成部分，既有理论价值又有实际应用. 同余是可除性的符号语言，在西方是由高斯最先引进的. 本章将讨论同余的概念和理论，并给出一些重要而有趣的应用.

在数论中一般只讨论正整数，但在讨论同余理论的时候我们把范围扩大，而讨论全体正负整数. 但当我们说到素数，除数，最大公约数时，仍然把它们看成正整数.

§7.1　同余式的性质

7.1.1　同余的定义

定义　给定一个正整数 m，把它叫做模. 如果用 m 去除整数 a 和 b 所得的余数相同，我们就说 a，b 对模 m 同余，记作

$$a \equiv b \pmod{m}; \tag{1}$$

如果余数不同，我们就说 a，b 对模 m 不同余，记为

$$a \not\equiv b \pmod{m}.$$

同余式(1) 读作"a 同余于 b，模 m". 同余式(1) 意味着 $m \mid (b-a)$，即

$$a - b = mk(k \text{ 是整数}).$$

例 1) $24 \equiv 9 \pmod 5$，因为 $24 - 9 = 15 = 5 \cdot 3$.

2) $47 \equiv 11 \pmod 9$，因为 $47 - 11 = 36 = 9 \cdot 4$.

3) $-11 \equiv 5 \pmod 8$，因为 $-11 - 5 = -16 = 8 \cdot (-2)$.

4) $81 \equiv 0 \pmod{27}$，$81 - 0 = 81 = 27 \cdot 3$

最后一例表明，在一般情况下我们可以用同余式

$$a \equiv 0 \pmod m$$

表明 m 能整除 a. 例如，一个数是偶数可以写为 $a \equiv 0 \pmod 2$. 类似地，一个数是奇数可以写为 $a \equiv 1 \pmod 2$.

同余的概念来自日常生活. 例如："星期三有一次课"，就含有同余的概念，所用的模是 7. 我国古代用的干支纪年也属于此类，其模为 60，每隔 60 年循环一次. 我国对同余式的研究有很光荣的历史，如孙子算经有"物不知其数"一题，这就是同余式研究的开始. 这一问题的原文如下：

"今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二，问物几何？"

这个问题同余的符号表示就是，求正整数 x，使得下式成立：

$$x \equiv 2 \pmod 3,$$
$$x \equiv 3 \pmod 5,$$
$$x \equiv 2 \pmod 7,$$

这三个同余方程的公共解就是问题的答案.

7.1.2 同余式的基本性质

同余式的记法使我们想起等式，事实上，同余式和等式有一些相同的性质，最简单的是以下三个：

1) 反身性， $a \equiv a \pmod m$；

2) 对称性， $a \equiv b \pmod m \Rightarrow b \equiv a \pmod m$；

3) 传递性， $a \equiv b \pmod m$ 和 $b \equiv c \pmod m \Rightarrow$
$a \equiv c \pmod m$.

3）的证明. 前两式意味着

$$a-b=mk, b-c=ml, \quad k, l \text{ 为整数}.$$

从而　$a-c=a-b+b-c=mk+ml=m(k+l)$.

所以　$a \equiv c \pmod m$.

例　由 $13 \equiv 35 \pmod{11}$，$35 \equiv -9 \pmod{11}$ 可推出

$$13 \equiv -9 \pmod{11}.$$

对于给定的数 b 和模 m，所有同余于 b 模 m 的数就是算术数列

$$b+km, \qquad k=0, \pm 1, \pm 2, \cdots.$$

有趣的是，通常的等式也可写为一种同余式，即模为 0 的同余式，这只要取

$$a \equiv b \pmod 0.$$

这时　$$a-b=0 \cdot k \text{ 或 } a=b.$$

这样，等式就作为特殊的同余式而出现. 但是在数学文献中很少使用这种表示.

以下我们总假定模 $m \geqslant 2$. 在同余式(1)中，若 $0 \leqslant b < m$，则称 b 是模 m 的最小非负剩余；若 $1 \leqslant b \leqslant m$，则称 b 是对模 m 的最小正剩余.

定理 1　a 同余于 b 模 m 的充要条件是 a 和 b 被 m 整除后所得的最小非负余数相等. 换言之，若

$$a=q_1 m+r_1, \quad 0 \leqslant r_1 < m$$

$$b=q_2 m+r_2, \quad 0 \leqslant r_2 < m$$

则 $r_1=r_2$.

证　由

$$a-b=(q_1-q_2)m+(r_1-r_2),$$

知道，$m \mid (a-b)$ 的充要条件是 $m \mid (r_1-r_2)$. 由于 $0 \leqslant |r_1-r_2| < m$，必有 $r_1-r_2=0$，即 $r_1=r_2$. 这样一来，任何整数 a 都可写为

$$a=km+r, \tag{2}$$

这里 k 是整数，r 是下面数中之一：$0,1,2,\cdots,m-1$.

例 1) $a=11, m=7$. $11=7\cdot1+4$

2) $a=-11, m=7$. $-11=7\cdot(-2)+3$.

7.1.3 同余式的四则运算

在研究同余式的运算时，自然地，我们将它们与代数中的等式类比．但要注意类比不是等同．我们既要找出共同点，又要找出不同点．共同点是，同余式可以相加、相减、相乘，这些运算与等式运算一样，也都成立．不同点表现在除法上．我们应给予特别的注意．

1) 若 $a\equiv b \pmod m$，$c\equiv d \pmod m$，则

$$a+c\equiv b+d \pmod m.$$

证 依定义，

$$a=b+mk, c=d+ml, \tag{3}$$

从而 $$a+c=b+d+m(k+l),$$

此式即 $$a+c\equiv b+d \pmod m.$$

同样的办法可以证明

2) 若 $a\equiv b \pmod m$，$c\equiv d \pmod m$，则

$$a-c\equiv b-d \pmod m.$$

例 $11\equiv-5 \pmod 8$，$7\equiv-9 \pmod 8$，

相加得 $$18\equiv-14 \pmod 8,$$

相减得 $$4\equiv4 \pmod 8.$$

3) 若 $a\equiv b \pmod m$，$c\equiv d \pmod m$，则 $ac\equiv bd \pmod m$.

证 由(3)

$$ac=(b+mk)(d+ml)=bd+m(kd+bl+mkl),$$

所以 $ac\equiv bd \pmod m$.

例 $11\equiv-5 \pmod 8$，$7\equiv-9 \pmod 8$．两式相乘得

$$77\equiv45 \pmod 8.$$

注 若 $a\equiv b \pmod m$，而 c 是任一整数，则

$$ac \equiv bc \pmod{m}.$$

这是性质 3) 中 $c = d$ 的特殊情况.

例 若 $11 \equiv -5 \pmod 8$,则 $33 \equiv -15 \pmod 8$. 这是同余式两边乘 3 得到的.

一个自然的问题是,在什么时候我们可以在同余式 $ac \equiv bc \pmod{m}$ 中消去公因数 c. 注意,在这里同余式的性质不同于等式.

例 $22 \equiv -2 \pmod 8$ 但约去 2 后,就得出 $11 \equiv -1 \pmod 8$,这是不对的. 这说明相乘是无条件的,相消就要加条件. 下面的 5),6),7) 提供了在什么条件下同余式两边因数可以相消.

4) 若 $a \equiv b \pmod{m}$,$k > 0$,则 $ak \equiv bk \pmod{mk}$.

证 $a \equiv b \pmod{m} \Rightarrow a - b = ml \Rightarrow ak - bk = mlk$

$\Rightarrow ka \equiv kb \pmod{km}$.

5) 若 $a \equiv b \pmod{m}$,d 是 a,b 和 m 的公因数,则

$$\frac{a}{d} \equiv \frac{b}{d} \left(\bmod \frac{m}{d}\right).$$

证 设则 $a = a_1 d$,$b = b_1 d$;$m = m_1 d$,则

$a \equiv b \pmod{m} \Rightarrow a - b = ml$,

$a_1 d - b_1 d = m_1 dl \Rightarrow a_1 - b_1 = m_1 l$

$\Rightarrow a_1 \equiv b_1 \pmod{m_1} \Leftrightarrow \frac{a}{d} \equiv \frac{b}{d} \left(\bmod \frac{m}{d}\right)$.

性质 4) 与 5) 构成一种对偶关系,即模与同余式可同乘同除一数.

例 考虑同余式 $33 \equiv 15 \pmod 9 \Leftrightarrow 3 \cdot 11 \equiv 3 \cdot 5 \pmod 9$. 由 $(3,9) = 3$,用 5) 得 $11 \equiv 5 \pmod 3$.

6) 若 $a \equiv b \pmod{m}$,$d \mid m$,$d > 0$,则 $a \equiv b \pmod{d}$.

证 设 $m = m_1 d$

$$a - b = ml = m_1 dl \Rightarrow a \equiv b \pmod{d}.$$

例 $14 \equiv 4 \pmod{10} \Rightarrow 14 \equiv 4 \pmod 5$;$14 \equiv 4 \equiv 0 \pmod 2$.

7）如果 $ca \equiv cb(\mathrm{mod}\ n)$，$d=(c,n)$ 是 c,n 的最大公约数，则 $a \equiv b(\mathrm{mod}\ \frac{n}{d})$.

证　根据假设，我们有

$$c(a-b)=ca-cb=kn(k\text{ 是整数}),$$

既然 $(c,n)=d$ 故可设 $c=dc_1, n=kn_1$. 所以

$$dc_1(a-b)=kdn_1 \Leftrightarrow c_1(a-b)=kn_1,$$

因此 $n_1 \mid c_1(a-b), (n_1,c_1)=1 \Rightarrow n_1 \mid (a-b) \Leftrightarrow a \equiv b(\mathrm{mod}\ n_1)$.

换言之，

$$a \equiv b(\mathrm{mod}\ \frac{n}{d}).$$

当 $(c,n)=1$ 时，直接从同余式两边消去 c，不必改变模.

系 1　若 $ca \equiv cb(\mathrm{mod}\ n)$，且 $(c,n)=1$，则 $a \equiv b(\mathrm{mod}\ n)$.

例　$-35 \equiv 45(\mathrm{mod}\ 8)$

$$5 \cdot (-7) \equiv 5 \cdot 9(\mathrm{mod}\ 8) \Rightarrow -7 \equiv 9(\mathrm{mod}\ 8).$$

系 1 的特殊情况是

系 2 若 $ca \equiv cb(\mathrm{mod}\ p)$，$p$ 是素数且 $p \nmid c$，则 $a \equiv b(\mathrm{mod}\ p)$.

例　$4 \equiv 48(\mathrm{mod}\ 11)$，约去 4，得 $1 \equiv 12(\mathrm{mod}\ 11)$.

7.1.4　同余式的方幂

作为同余式乘法的推论，我们可以得到关于同余式的方幂.

8）若 $a \equiv b(\mathrm{mod}\ m)$，则 $a^n \equiv b^n(\mathrm{mod}\ m)$.

这一结果可以用来求一个数的高次幂被某数除的余数.

例　$3^{89} \equiv ?(\mathrm{mod}\ 7)$，

解　$9=3^2 \equiv 2(\mathrm{mod}\ 7)$，　　$3^4 \equiv 4(\mathrm{mod}\ 7)$，

$3^8 \equiv 16 \equiv 2(\mathrm{mod}\ 7)$，　　$3^{16} \equiv 4(\mathrm{mod}\ 7)$，

$3^{32} \equiv 16 \equiv 2(\mathrm{mod}\ 7)$，　　$3^{64} \equiv 4(\mathrm{mod}\ 7)$.

而 $89=64+16+8+1$，由此推出

$$3^{89}=3^{64} \cdot 3^8 \cdot 3 \equiv 4 \cdot 4 \cdot 2 \cdot 3=96 \equiv 5(\mathrm{mod}\ 7).$$

利用同余式的性质，我们很容易地求出了 3^{89} 的同余数(mod 7). 要是

把它乘出来算要花很大力气.

一个有趣的应用是关于费马数的,以

$$F_t = 2^{2^t} + 1$$

表示费马数,则前四个费马数是

$$F_0 = 3, F_1 = 5, F_2 = 17, F_3 = 257, F_4 = 65537.$$

除了 F_0 和 F_1 外,所有的费马数均以 7 结尾.我们利用同余式证明,结果正是这样.显然,这等价于

$$F_t - 1 = 2^{2^t}, t = 2, 3, \cdots$$

都以 6 结尾,即它们被 10 除都余 6.我们用归纳法来证明.当 $t = 2$ 时,结果显然是对的.今假定对某个 t,

$$2^{2^t} \equiv 6 \pmod{10},$$

两边取平方得到

$$2^{2^{t+1}} \equiv 36 \equiv 6 \pmod{10}.$$

这就是要证明的.

定理 2 若 $P(x) = \sum_{k=0}^{m} c_k x^k$ 是一个多项式,系数 c_k 是整数.如果 $a \equiv b \pmod{n}$,则 $P(a) \equiv P(b) \pmod{n}$.

证 由 $a \equiv b \pmod{n}$ 可得 $a^k \equiv b^k \pmod{n}$, $k = 0, 1, 2, \cdots, m$.因此 $c_k a^k \equiv c_k b^k \pmod{n}$.

对于所有的 k,把这 $m + 1$ 个同余式加起来,我们得到

$$\sum_{k=0}^{m} c_k a^k \equiv \sum_{k=0}^{m} c_k b^k \pmod{n}.$$

也就是 $P(a) \equiv P(b) \pmod{n}$.

如果 $P(x)$ 是一个整系数多项式,而且 $P(a) \equiv 0 \pmod{n}$,我们就说 a 是同余式 $P(x) \equiv 0 \pmod{n}$ 的一个解.

系 若 a 是 $P(x) \equiv 0 \pmod{n}$ 的一个解,且 $a \equiv b \pmod{n}$,则 b 也是它的一个解.

证　由上述定理，可知 $P(a)\equiv P(b)\pmod n$. 那么，如果 a 是 $P(x)\equiv 0\pmod n$ 的一个解，则 $P(a)\equiv P(b)\equiv 0\pmod n$，也就是 b 是此式的一个解.

另一个有趣的应用是关于偶完全数的. 我们在前面指出过，偶完全数 n 都以 6 和 8 结尾. 现在给予证明.

定理 3　偶完全数 n 总以 6 和 8 结尾，即 $n\equiv 6\pmod{10}$ 或 $n\equiv 8\pmod{10}$.

证　一个偶完全数总可以表示为(见第 4 章 §1，定理 6)

$$n = 2^{k-1}(2^k - 1),$$

其中 $2^k - 1$ 是素数. 前面已经证明 k 一定是素数. 如果 $k = 2$，则 $n = 6$，正是定理所说的. 所以我们将注意力集中于 $k > 2$ 的情形. 任一奇素数，或者被 4 除余 1，或者被 4 除余 3. 下面依 k 取 $4m+1$ 的形式，或 k 取 $4m+3$ 的形式，分成两种情形证明.

若 k 是 $4m+1$ 的形式，则

$$n = 2^{4m}(2^{4m+1} - 1) = 2^{8m+1} - 2^{4m} = 2\cdot 16^{2m} - 16^m$$

借助简单的归纳法，容易证明，对任意的 t，都有

$$16^t \equiv 6\pmod{10}$$

利用这一同余式可得，

$$n = 2\cdot 16^{2m} - 16^m \equiv 2\cdot 6 - 6 \equiv 6\pmod{10}.$$

其次，考虑 $k = 4m+3$ 的情况. 这时，

$$n = 2^{4m+2}(2^{4m+3} - 1) = 2^{8m+5} - 2^{4m+2} = 2\cdot 16^{2m+1} - 4\cdot 16^m.$$

记住 $16^t \equiv 6\pmod{10}$，我们得到：

$$n \equiv 2\cdot 6 - 4\cdot 6 \equiv -12 \equiv 8\pmod{10}.$$

也就是每个偶完全数都是以 6 或 8 结尾的.

7.1.5　检查因数的方法.

作为同余式的应用，这里给出两个检查因数的方法，并予以证明.

定理 4　一个整数能被 3 或 9 整除的充要条件是它的十进位数

码的和能被 3 或 9 整除.

证 设 a 是任一正整数.把 a 写成十进位数的形式:

$$a=a_n10^n+a_{n-1}10^{n-1}+\cdots+a_0, 0\leqslant a_i<10, i=0,1,\cdots,n.$$

因为 $10\equiv1 \pmod 3$,所以

$$a\equiv a_n+a_{n-1}+\cdots+a_0 \pmod 3.$$

由此可知,$3\mid a$ 当且仅当 $3\mid(a_n+a_{n-1}+\cdots+a_0)$.同法可证 9 的情况.

例 若 $a=5874192$,则由定理 4

$$\sum_{i=0}^{n}a_i=5+8+7+4+1+9+2=36.$$

36 能被 3,9 整除.故 a 能被 3,9 整除.

例 若 $a=435693$,则由定理 4,

$$\sum_{i=0}^{n}a_i=4+3+5+6+9+3=30.$$

30 能被 3 整除,故 3 是 a 的因数.但 $\sum\limits_{i=0}^{n}a_i$ 不能被 9 整除,故 9 不是 a 的因数.

正整数 a 可表示为

$$a=a_n1000^n+a_{n-1}1000^{n-1}+\cdots+a_0,$$
$$0\leqslant a_i<1000, i=0,1\cdots,n.$$

由此,得下面的定理.

定理 5 a 能被 7(或 11,或 13)整除的充要条件是 7(或 11,或 13)整除

$$(a_0+a_2+\cdots)-(a_1+a_3+\cdots)=\sum_{i=0}^{n}(-1)^i a_i.$$

证 通过直接计算可知,$1000\equiv-1 \pmod 7$.从而

$$1000^2\equiv1,\ 1000^3\equiv-1,\cdots,1000^n\equiv(-1)^n \pmod 7,$$

所以

$$a \equiv a_n(-1)^n + a_{n-1}(-1)^{n-1} + \cdots - a_1 + a_0 \pmod 7$$

$$(a_0 + a_2 + \cdots) - (a_1 + a_3 + \cdots) = \sum_{i=0}^{n}(-1)^i a_i.$$

因为 $1000 \equiv -1 \pmod{11}$ 和 $1000 \equiv -1 \pmod{13}$，所以同样的推理对模 11 和模 13 也成立. 定理证毕.

例 若 $a = 637693$，则 $a = 637 \cdot 1000 + 693$，

$$\sum_{i=0}^{n}(-1)^i a_i = 693 - 637 = 56$$

能被 7 整除而不能被 11 与 13 整除. 故由定理 5，7 是 a 的因数. 但 11，13 不是 a 的因数.

例 若 $a = 75312289$，则 $a = 75 \cdot 1000^2 + 312 \cdot 1000 + 289$，

$$\sum_{i=0}^{n} a_i = 289 - 312 + 75 = 52$$

能被 13 整除，而不能被 7，11 整除. 故由定理 5，13 是 a 的因数，而 7 与 11 不是 a 的因数.

7.1.6 弃九法(验算整数计算结果的方法)

这里讲的方法对加、减和乘都是正确的，且证法一样，所以只给出乘的证明. 假设我们由普通乘法运算求出整数 a，b 的乘积是 P，并令

$$a = a_n 10^n + a_{n-1}10^{n-1} + \cdots + a_0, \quad 0 \leqslant a_i < 10, i = 0,1,\cdots,n,$$

$$b = b_m 10^m + b_{m-1}10^{m-1} + \cdots + b_0, \quad 0 \leqslant b_j < 10, j = 0,1,\cdots,m,$$

$$P = c_l 10^l + c_{l-1}10^{l-1} + \cdots + c_0, \quad 0 \leqslant c_k < 10, k = 0,1,\cdots,l.$$

我们说：如果

$$\left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right) \not\equiv \sum_{k=0}^{l} c_k \pmod 9, \tag{4}$$

那么所求得的乘积是错误的. 利用前面的性质，可以很容易的证明它.

$10 \equiv 1(\bmod 9) \Rightarrow 10^n \equiv 1(\bmod 9)$. 由此,

$$a = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_0$$
$$\equiv a_n + a_{n-1} + \cdots + a_0(\bmod 9),$$
$$b = b_m 10^m + b_{m-1} 10^{m-1} + \cdots + b_0$$
$$\equiv b_m + b_{m-1} + \cdots + b_0(\bmod 9),$$

从而 $\quad ab \equiv \left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right)(\bmod 9), P \equiv \sum_{k=0}^{l} c_k(\bmod 9).$

必有

$$\left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right) \equiv \sum_{k=0}^{l} c_k(\bmod 9),$$

所以,若 $ab \not\equiv P(\bmod 9)$,则 ab 不是 P.

以上所说就是弃九法的原理.在实际验算时,若 a_i, b_j, c_k 中有 9 出现,还可以去掉(因 $9 \equiv 0(\bmod 9)$).我们看一个例子.

例　设 $a = 28997, b = 39495$. 如果按照普通计算方法得到 a, b 的乘积是 $P = 1145236415$,那么我们按照上述方法

$$a \equiv 17(\bmod 9), b \equiv 3(\bmod 9), P = 32(\bmod 9).$$

但

$$3 \cdot 17 \not\equiv 32(\bmod 9)$$

故知计算有误.

依照上述方法的道理,同样可以得出验算和,差的正确性的方法.这个验算方法的优点在于很容易求出(4)式,因此验算可以进行得比较快.

但是应该特别注意,当使用弃九法时,得出的结果虽然是

$$\left(\sum_{i=0}^{n} a_i\right)\left(\sum_{j=0}^{m} b_j\right) \equiv \sum_{k=0}^{l} c_k(\bmod 9),$$

也还不能完全肯定原计算是正确的.例如在上面的例中,正确的结果是 1145236515.如果有人计算出来的结果是 1145235615.那么用弃九法,就得

$$3 \cdot 17 \equiv 33(\bmod 9),$$

而并未检查出错误来,因此这个验算方法是有它的缺点的.

7.1.7　剩余类与完全剩余系

有了同余的概念,我们就可以把余数相同的数放在一起,这样就产生了剩余类的概念.若 m 是一个给定的正整数,则全部整数可以分成 m 个集合,记作 $K_0,K_1,\cdots,K_{m-1}$,其中 $K_r(r=0,1,\cdots,m-1)$ 是由一切形如 $qm+r(q=0,\pm1,\pm2,\cdots)$ 的整数所组成.从前面的讨论不难看出,这些集合具有如下两条性质:

1) 每一个整数必包含在上述的一个集合之中,且仅包含在一个之中.

2) 两个整数在同一个集合中的充要条件是这两个整数对模 m 同余.

因此我们引出剩余类和完全剩余系的概念.

定义　$K_0,K_1,\cdots,K_{m-1}$ 叫做模 m 的剩余类.若 $a_0,a_1,\cdots,a_{m-1}$ 是 m 个整数,并且其中任何两个都不在同一个剩余类中,则 $a_0,a_1,\cdots,a_{m-1}$ 叫做模 m 的一个完全剩余系.

例　取 $m=3$.则模 3 有三个剩余类:

$K_0=\{0,3,6,9,\cdots\},K_1=\{1,4,7,10,\cdots\},K_2=\{2,5,8,11,\cdots\}.$

容易看出,0,1,2 构成模 3 的一个完全剩余系.类似地,9,10,11 也构成模 3 的一个完全剩余系.

当 m 是偶数时,序列

$$-\frac{m}{2},-\frac{m}{2}+1,\cdots,-1,0,1,\cdots,\frac{m}{2}-1,$$

$$\frac{m}{2}+1,\cdots,-1,0,1\cdots,\frac{m}{2}-1,\frac{m}{2}$$

都是模 m 的完全剩余系.

当 m 是奇数时,序列

$$-\frac{m-1}{2},\cdots,-1,0,1,\cdots,\frac{m-1}{2}$$

是 m 的完全剩余系.

定义 0,1,…,$m-1$ 这 m 个整数叫做模 m 的最小非负完全剩余系;当 m 是偶数时, $\frac{m}{2}$,…,-1,0,1,…,$\frac{m}{2}-1$ 或 $\frac{m}{2}+1$,…,-1,0,1,…,$\frac{m}{2}$ 叫做模 m 的绝对最小完全剩余系;当 m 是奇数时,$-\frac{m-1}{2}$,…,-1,0,1,…,$\frac{m-1}{2}$ 叫做模 m 的绝对最小完全剩余系.

定理 6 如果 x 取模 m 的一个完全剩余系的所有值,并且 $(a,m)=1$,那么 $ax+b$ 也可取一个完全剩余系的所有值.

证 我们只须证明

$x_1 \equiv x_2 \Rightarrow ax_1+b \equiv ax_2+b$. $x_1 \not\equiv x_2 \Rightarrow ax_1+b \not\equiv ax_2+b$

这是明显的.

定理 7 若 m,n 是互素的两个整数,x,y 分别通过 m,n 的完全剩余系,则 $nx+my$ 通过 $m\cdot n$ 的完全剩余系.

证 由假设知道,x,y 分别通过 m,n 个整数,因此 $nx+my$ 分别通过 $m\cdot n$ 个整数.因此我们只需证明这个 $m\cdot n$ 整数对模 $m\cdot n$ 两两不同余就够了.假定

$$nx_1+my_1 \equiv nx_2+my_2 \pmod{mn},$$

其中 x_1,x_2 是 x 所通过的完全剩余系的整数,而 y_1,y_2 是 y 所通过的完全剩余系的整数,不妨设这两个完全剩余系是最小非负完全剩余系,则由定理 6)

$$nx_1+my_1 \equiv nx_2+my_2 \pmod{m} \tag{5}$$

$$\Rightarrow nx_1 \equiv nx_2 \pmod{m},\ (m,n)=1 \Rightarrow x_1 \equiv x_2 \pmod{m}.$$

同理 $nx_1+my_1 \equiv nx_2+my_2 \pmod{n}$

$$\Rightarrow my_1 \equiv my_2 \pmod{n},\ (m,n)=1 \Rightarrow y_1 \equiv y_2 \pmod{n}.$$

若 x_1,x_2 都在同一个完全剩余系中,则 $x_1=x_2$.

若 y_1,y_2 都在同一个完全剩余系中,则 $y_1=y_2$.

这表明若 x_1, y_1 与 x_2, y_2 中有一个不同，则同余式(5)不会成立.

习　题

1. 证明任何整数(十进位表示)模 9 同余于它的各位数字之和.

2. 设正整数

$$a = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_0,$$

试证 11 整除 a 的充要条件是

$$11 \mid \sum_{i=0}^{n} (-1)^i a_i.$$

3. 用检查因式的方法求 1535625, 1158066 的素因子分解式.

4. 解释下面这一组有趣的等式.

$$1 \cdot 9 + 2 = 11,$$
$$12 \cdot 9 + 3 = 111,$$
$$123 \cdot 9 + 4 = 1111,$$
$$1234 \cdot 9 + 5 = 11111,$$
$$12345 \cdot 9 + 6 = 111111,$$
$$123456 \cdot 9 + 7 = 1111111,$$
$$1234567 \cdot 9 + 8 = 11111111,$$
$$12345678 \cdot 9 + 9 = 111111111,$$
$$123456789 \cdot 9 + 10 = 1111111111$$

(提示：证明等式 $(10^{n-1} + 2 \cdot 10^{n-2} + 3 \cdot 10^{n-3} + \cdots + n)(10-1) + (n+1) = (10^{n+1} - 1)/9$.)

§7.2 中国剩余定理

7.2.1 同余式

在代数里一个主要的问题是解代数方程. 现在我们讨论与解方程类似的问题:求同余式的解. 我们先讨论一次同余式,再讨论同余式组. 我们还要特别介绍一下中国古代数学家在这方面的卓越成就. 为此先引进基本概念.

定义 设 $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$,其中 a_i 是整数;又设 m 是一个正整数,则

$$f(x) \equiv 0 \pmod{m} \tag{1}$$

叫做模 m 的同余式. 若 $a_n \not\equiv 0 \pmod{m}$ 则 n 叫做同余式(1)的次数.

由前面的讨论知,若 $f(a) \equiv 0 \pmod{m}$,且 $a \equiv a' \pmod{m}$,则 $f(a') \equiv 0 \pmod{m}$. 因此我们引入如下的定义.

定义 若 a 是使 $f(a) \equiv 0 \pmod{m}$ 成立的一个整数,则 $x \equiv a \pmod{m}$ 叫做(1)的一解. 即,满足(1)而对模 m 相互同余的一切数算作(1)的一个解.

形如
$$ax \equiv b \pmod{n}$$
的同余式是一次同余式. 若 x_0 是它的一个解,则

$$ax_0 \equiv b \pmod{n} \Leftrightarrow n \mid (ax_0 - b) \Leftrightarrow \exists\, y_0, 使\ ax_0 - b = ny_0.$$

这样一来,求一次同余式解的问题等价于解一个二元一次不定方程. 这就容许我们使用前面的结果了(参考第 5 章).

定理 1 一次同余式 $ax \equiv b \pmod{n}$ 有解的充要条件是 $d \mid b$,这里 $d = (a, n)$. 如果 $d \mid b$,则它有 d 个模 m 彼此不同余的解.

证 证明分为两步:1) 存在 d 个彼此不同余的解. 证明的方法是构造性的,具体地把 d 个解写出来;2) 解不多于 d 个,即任一解必属于上述 d 个解之一.

1）我们已经知道，给定的同余式等价于不定方程

$$ax - ny = b.$$

它可解的充要条件是 $d \mid b$. 如果它可解，且 x_0, y_0 是它的一个特解，则它的一般解为

$$x = x_0 + \frac{n}{d}t, y = y_0 + \frac{a}{d}t.$$

令 $t = 0,1,2,\cdots,d-1$，考虑

$$x_0, x_0 + \frac{n}{d}, x_0 + 2\frac{n}{d}, \cdots, x_0 + (d-1)\frac{n}{d}. \tag{2}$$

我们指出，这些整数是模 n 彼此不同余的，而其它整数 x 同余于它们中的一个. 事实上，如果

$$x_0 + \frac{n}{d}t_1 \equiv x_0 + \frac{n}{d}t_2 \pmod{n}, 0 \leqslant t_1 < t_2 \leqslant d-1,$$

则

$$\frac{n}{d}t_1 \equiv \frac{n}{d}t_2 \pmod{n}.$$

现在 $(n/d, n) = n/d$，所以根据同余式的性质，因子 n/d 可以消去，而得到

$$t_1 \equiv t_2 \pmod{d},$$

从而

$$d \mid (t_2 - t_1) \Rightarrow t_1 = t_2.$$

2）剩下来要证明的是，任何一个其它解 $x_0 + (n/d)t$ 一定模 n 同余于(2)中罗列的 d 个整数中的一个. 把 t 表示为 $t = qd + r, 0 \leqslant r \leqslant d-1$. 我们有，

$$x_0 + \frac{n}{d}t = x_0 + \frac{n}{d}(qd + r)$$

$$= x_0 + nq + \frac{n}{d}r \equiv x_0 + \frac{n}{d}r \pmod{n}$$

这是(2)中之一. 证毕.

定理 1 不但给出了可解的充要条件，而且在可解的条件下，给出了解的表达式. 更明确些，若 x_0 是 $ax \equiv b \pmod{n}$ 的任一个特解，则

它的 $d=(a,n)$ 个彼此不同余的解是

$$x_0, x_0+n/d, x_0+2(n/d), \cdots, x_0+(d-1)(n/d).$$

特别地，当 a 和 n 互素时，我们有，

系 若 $(a,n)=1$，则同余式 $ax\equiv b(\mathrm{mod}\ n)$ 有唯一解.

例 求解 $18x\equiv 30(\mathrm{mod}\ 42)$.

解 因为 $(18,42)=6$，且 6 整除 30，所以同余式有解，且有 6 个彼此不同余的解. 由视察法，$x=4$ 是一个解. 由定理 1，6 个解是

$$x\equiv 4+(42/6)t\equiv 4+7t(\mathrm{mod}\ 42),\quad t=0,1,\cdots,5.$$

具体算出来，它们是

$$x\equiv 4,11,18,25,32,39(\mathrm{mod}\ 42).$$

例 求解 $9x\equiv 21(\mathrm{mod}\ 30)$.

解 因为 $(9,30)=3$ 和 $3\mid 21$，所以我们知道，同余式有解，且有 3 个彼此不同余的解. 现在换一种解法来解. 用 3 除同余式的两边，得

$$3x\equiv 7(\mathrm{mod}\ 10).$$

由于 $(3,10)=1$，所以这个同余式，模 10 有唯一解. 从 $0,1,2,\cdots,9$ 依次作检验，可以把解求出来. 这个方法当然有点笨. 一个好一点的方法是在同余式的两边乘 7，得到

$$21x\equiv 49(\mathrm{mod}\ 10)\Leftrightarrow x\equiv 9(\mathrm{mod}\ 10).$$

（这一简化不是碰巧如此，因为 $0\times 3, 1\times 3, \cdots, 9\times 3$ 形成模 10 的完全剩余，所以其中必有一个模 10 同余于 1.）我们用这个办法找到了一个特解，但是原来的同余式中模是 30，所以要在 $0,1,2,\cdots,29$ 中找不同余的解. 在公式

$$x=9+10t$$

中取 $t=0,1,2$，得到 9，19，29. 回到原解，我们有

$$x\equiv 9(\mathrm{mod}\ 30), x\equiv 19(\mathrm{mod}\ 30), x\equiv 29(\mathrm{mod}\ 30).$$

这就是问题的解.

另一个求特解的不同方法来自定理 1 的证明. 同余式等价于一个不定方程：

$$9x \equiv 21 \pmod{30} \Leftrightarrow 9x - 30y = 21.$$

为了解不定方程，把 $3 = (9,30)$ 表示为 9 与 30 的线性组合：$3 = 9 \cdot (-3) + 30 \cdot 1$. 所以

$$21 = 7 \cdot 3 = 9 \cdot (-21) - 30 \cdot (-7).$$

这样一来，$x = -21, y = -7$ 满足不定方程. 因此，同余式的所有解都可以从公式

$$x = -21 + 10t$$

中找到. 令 $t = 0,1,2$ 是模 30 不同余的解(但是它们对模 10 都同余). 这样一来，我们得到彼此不同余的解：

$$x \equiv -21 \pmod{30}, x \equiv -11 \pmod{30}, x \equiv -1 \pmod{30}.$$

把它们化为正整数，就得到 $x \equiv 9,19,29 \pmod{30}$.

7.2.2 中国剩余定理

上节讨论了一个未知数的同余式的解法，本节要讨论如何解下面重要的同余式组

$$x \equiv b_1 \pmod{m_1}, x \equiv b_2 \pmod{m_2}, \cdots, x \equiv b_k \pmod{m_k}.$$

在我国古代的孙子算经(纪元前后)里已经提出了这种形式的问题，这就是前面提出的物不知其数的问题. 孙子算经的答案是："答曰二十三".

孙子算经是中国古代的优秀著作，但作者和出版年代已无法考证了. 有人说，这是孙武的作品(即写孙子兵法 13 篇的作者). 但也有人反对. 有人根据其中的内容判断，认为是汉魏时的作品. 例如清代的戴震就根据书中涉及到"长安到洛阳的距离"和"佛书二十九章"等语，判定作者是汉明帝以后的人. 又如，清代阮元根据其中棋局十九道，而断定为汉以后的人. 但有人反对这个意见，因为古代在一本名著重新刊印的时候，常常掺杂了后人的补充材料. 因此著作年代还不能确定，连哪个世纪也不能确定，但在战国到三国之间不会错. 尽管如此，它仍然是我国最古老的三大数学名著之一. 这三大名著是，

周髀算经，九章算术，孙子算经. 特别是，“物不知其数”一题是世界上公认的最古老的重要工作.

定理2 设 $n_1, n_2, \cdots, n_r$ 是正整数，并且 $(n_i, n_j) = 1, i \neq j$. 那么一元同余式组

$$x \equiv a_1 \pmod{n_1}, x \equiv a_2 \pmod{n_2}, \cdots, x \equiv a_r \pmod{n_r}$$

有一个模 $n_1 n_2 \cdots n_r$ 的公共解.

证 证明分为两步：1）证明解存在；这一步是构造性的，即造一个解；2）证明解是唯一的；若另有一解，则它同余于上述的解.

1）令 $n = n_1 n_2 \cdots n_r$. 对于每一个 $k = 1, 2, \cdots, r$，令

$$N_k = n/n_k = n_1 \cdots n_{k-1} n_{k+1} \cdots n_r.$$

换句话说，N_k 是除去 n_k 的所有整数 n_i 的乘积. 由假设，n_i 是两两互素的，所以 $(N_k, n_k) = 1$. 根据一元同余式中的定理，同余式 $N_k x \equiv 1 \pmod{n_k}$ 是可解的，设它的唯一解为 x_k. 令

$$\overline{x} = a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_r N_r x_r.$$

我们的目的是证明整数 $\overline{x}$ 是所给同余式组的一个公共解.

首先. 通过观察可知 $N_j \equiv 0 \pmod{n_k}, j \neq k$，因为 $n_k | N_j$. 那么

$$\overline{x} = a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_r N_r x_r \equiv a_k N_k x_k \pmod{n_k}.$$

又整数 x_k 满足同余式 $N_k x \equiv 1 \pmod{n_k}$，则

$$\overline{x} \equiv a_k \cdot 1 \equiv a_k \pmod{n_k}.$$

这说明上述同余式组存在一个解.

2）下面证明解的唯一性，假设 x' 是另一个满足上述同余式组的解，则

$$\overline{x} \equiv a_k \cdot 1 \equiv x' \pmod{n_k}, \qquad k = 1, 2, \cdots, r.$$

并且对于每一个 k 有 $n_k | (\overline{x} - x')$. 因为 $(n_i, n_j) = 1$，所以，$n_1 n_2 \cdots n_r | (\overline{x} - x')$，于是 $\overline{x} \equiv x' \pmod{n}$. 证毕.

注 在中国古书中，运算的每一步都有称呼：n 叫最小公倍数，N_k 叫衍数，x_k 叫乘率. 因而解法步骤可总结如下：

1. 求最小公倍数和衍数：

$$最小公倍数\ n = n_1 n_2 \cdots n_r.$$

$$衍数：\begin{cases} N_1 = n_2 n_3 \cdots n_r = \dfrac{n}{n_1}, \\ N_2 = n_1 n_3 \cdots n_r = \dfrac{n}{n_2}, \\ \cdots\cdots\cdots\cdots \\ N_r = n_1 n_2 \cdots n_{r-1} = \dfrac{n}{n_r}. \end{cases}$$

2. 解同余式求乘率：

$$N_1 x \equiv 1 (\mathrm{mod}\ n_1) \Rightarrow x_1,$$

$$N_2 x \equiv 1 (\mathrm{mod}\ n_2) \Rightarrow x_2,$$

$$\cdots\cdots\cdots\cdots$$

$$N_r x \equiv 1 (\mathrm{mod}\ n_r) \Rightarrow x_r.$$

3. 构造解：$x = a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_r N_r x_r$.

例 首先我们解决前面提到的《孙子算经》中的同余式组问题.

我们已经在前面列出了同余式组：

$$x \equiv 2(\mathrm{mod}\ 3), x \equiv 3(\mathrm{mod}\ 5), x \equiv 2(\mathrm{mod}\ 7).$$

由定理 2，可知 $n = 3 \cdot 7 \cdot 5 = 105, N_1 = n/3 = 35, N_2 = n/5 = 21, N_3 = n/7 = 15$，则一元同余式

$$35x \equiv 1(\mathrm{mod}\ 3), 21x \equiv 1(\mathrm{mod}\ 5), 15x \equiv 1(\mathrm{mod}\ 7)$$

分别有解 $x_1 = 2, x_2 = 1, x_3 = 1$. 这样同余式组的一个解为

$$\overline{x} = 2 \cdot 35 \cdot 2 + 3 \cdot 21 \cdot 1 + 2 \cdot 15 \cdot 1 = 233.$$

取 105 为模，我们得到同余式组的唯一解 $\overline{x} = 233 \equiv 23(\mathrm{mod}\ 105)$.

孙子算经里面所用的方法可以列表如下：

除数	余数	最小公倍数	衍数	乘率	各总	答数	最小答案
3	2	$3\times5\times7=105$	5×7	2	$35\times2\times2$	233	23
5	3		7×3	1	$21\times1\times3$		
7	2		3×5	1	$15\times1\times2$		

例 解一元同余式

$$17x \equiv 9(\mathrm{mod}\ 276).$$

由 $276 = 3\cdot4\cdot23$,这等价于解同余式组

$$17x \equiv 9(\mathrm{mod}\ 3), 17x \equiv 9(\mathrm{mod}\ 4), 17x \equiv 9(\mathrm{mod}\ 23).$$

或 $x \equiv 0(\mathrm{mod}\ 3), x \equiv 1(\mathrm{mod}\ 4), 17x \equiv 9(\mathrm{mod}\ 23).$

注意如果 $x \equiv 0(\mathrm{mod}\ 3)$,则 $x = 3k$. 对于每一个 k,把 $3k$ 代入第二个同余式得到

$$3k \equiv 1(\mathrm{mod}\ 4).$$

两边同时乘以 3

$$9k \equiv 3(\mathrm{mod}\ 4),又\ 9k \equiv k(\mathrm{mod}\ 4) \Rightarrow k \equiv 3(\mathrm{mod}\ 4)$$

也就是 $k = 3 + 4j$,这里 j 是正整数. 于是

$$x = 3(3+4j) = 9 + 12j.$$

要 x 满足最后一个同余式,必须有

$$17(9+12j) \equiv 9(\mathrm{mod}\ 23),$$

或 $204j \equiv -144(\mathrm{mod}\ 23)$. 此式可化简为 $j \equiv 2(\mathrm{mod}\ 23)$,从而 $j = 2 + 23t$, t 取整数. 由此

$$x = 9 + 12j = 9 + 12(2+23t) = 33 + 276t.$$

即 $x \equiv 33(\mathrm{mod}\ 276)$. 这就是上述同余式组的解,也就是 $17x \equiv 9(\mathrm{mod}\ 276)$ 的解.

7.2.3 程大位的口诀

程大位在"算法统宗"(1592)中以诗的语言写出了孙子问题的算法口诀:

① 三人同行七十稀,② 五树梅花廿一枝,

③ 七子团圆月正半,④ 除百零五便得知.

程大位生于明嘉靖十二年四月初十(公元1533年5月3日)卒于万历三十四年八月十七日(公元1609年9月18日).他的"算法统宗"传入日本、朝鲜及东南亚,对那里的数学发展有很大的影响.现在看口诀的含义.

① 用70乘被3除的余数:$70\times 2=140$,

② 用21乘被5除的余数:$21\times 3=63$,

③ 用15乘被7除的余数:$15\times 2=30$.

然后加起来 $70\times 2+21\times 3+15\times 2=233$.

④$233-105-105=23$.

为什么70,21,15有如此妙用?

70:被3除余1,而被5,7除尽,

21:被5除余1,而被3,7除尽,

15:被7除余1,而被3,5除尽.

所以,

$70a$:被3除余 a,而被5,7除尽,

$21b$:被5除余 b,而被3,7除尽,

$15c$:被7除余 c,而被3,5除尽.

这样一来,$70a+21b+15c$ 被3除余 a,被5除余 b,被7除余 c.

程大位的口诀里,前三句的意义:点出3,5,7与70,15,21的关系.后一句指出求最小正解还需减105.

这个方法的原则反映在插值理论,代数理论及算子理论中.今举一例:拉格朗日插值法.

问题　找到一个函数,在 a,b,c 三点取 α,β,γ 三值.

孙子方法给我们提供了解决问题的途径:

1. 作函数 $p(x)$,使 $p(a)=1,p(b)=p(c)=0$.

   作函数 $q(x)$,使 $q(a)=1,q(b)=q(c)=0$.

   作函数 $r(x)$,使 $r(c)=1,r(a)=r(b)=0$.

2. $\alpha p(x)+\beta q(x)+\gamma r(x)$ 满足条件.

可见,解法简单明白.只要求出 $p(x),q(x),r(x)$ 答案就出来了.它们好求吗?很好求.具体求法是:取 $p(x)=\lambda(x-b)(x-c)$,

令
$$p(x)=1\Rightarrow p(a)=\lambda(a-b)(a-c)=1$$
$$\Rightarrow\lambda=\frac{1}{(a-b)(a-c)}.$$
所以
$$P(x)=\frac{(x-b)(x-c)}{(a-b)(a-c)}.$$
同理可得
$$q(x)=\frac{(x-c)(x-a)}{(b-c)(b-a)},r(x)=\frac{(x-a)(x-b)}{(c-a)(c-b)}.$$
因此
$$\alpha\frac{(x-b)(x-c)}{(a-b)(a-c)}+\beta\frac{(x-c)(x-a)}{(b-c)(b-a)}+\gamma\frac{(x-a)(x-b)}{(c-a)(c-b)}$$
就是问题的一个答案.这就是著名的插值法中的拉格朗日公式.从孙子的原则来看,推导是多么简单明了.

数学在应用的时候,一般仅仅有有限个数据,我们就用这一类的方法来推演出函数来,然后用它描述其他各点的大概数据.

在 n 个不同点 $a_1,\cdots,a_n$ 函数 $f(x)$ 各取值 $\alpha_1,\cdots,\alpha_n$ 的插值公式是
$$\alpha_1\frac{(x-a_2)\cdots(x-a_n)}{(a_1-a_2)\cdots(a_1-a_n)}+\alpha_2\frac{(x-a_1)(x-a_3)\cdots(x-a_n)}{(a_2-a_1)(a_2-a_3)\cdots(a_2-a_n)}$$
$$+\cdots+\alpha_n\frac{(x-a_1)\cdots(x-a_{n-1})}{(a_n-a_1)\cdots(a_n-a_{n-1})},$$
这是不必证明的公式了!

由此看来"插值公式"与"70,21,15"法,面貌虽不同,原则本无隔.在孙子算法中可以差一个105倍数,而这里可以差一个在 $a_1,\cdots,a_n$ 点都等于0的函数.这个方法提供了以下的原则.

要作出一个具有性质 A,B,C 的数学结构.我们的办法是先作

出"单因子构件".也就是作出性质 B,C 不发生作用,而性质 A 取单位量的构件.再作出性质 C,A 不发生作用,而性质 B 取单位量的构件.最后作出性质 A,B 不发生作用,而性质 C 取单位量的构件.那么所要求的构件可以由这些构件凑出来.这种方法在高等数学中经常碰到.

习　题

1. 求下列各同余式的解:

(i) $256x \equiv 179 \pmod{337}$,　　(ii) $1215x \equiv 560 \pmod{2755}$,

(iii) $1296x \equiv 1125 \pmod{1935}$.

2. 韩信点兵:有兵一队,若列成五行纵队,则末行一人,列成六行纵队,则末行五人,列成七行纵队,则末行四人,列成十一行纵队,则末行十人,求兵数.

3. 试解下列各题:(i) 十一数余三,七二数余二,十三数余一,问本数.

(ii) 二数余一,五数余二,七数余三,九数余四,问本数.

(杨辉:续古摘奇算法(1275)).

§7.3　费马小定理与欧拉定理

7.3.1　费马小定理

从二项式定理开始.我们有

$$(x+y)^2 = x^2 + 2xy + y^2,$$
$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3.$$

在一般情况下,

$$(x+y)^p = x^p + C_p^1 x^{p-1} y + C_p^2 x^{p-2} y^2 + \cdots + C_p^{p-1} x y^{p-1} + y^p \qquad (1)$$

中间的二项式系数是

$$C_p^1 = \frac{p}{1}, \qquad C_p^2 = \frac{p(p-1)}{2!},$$

$$C_p^3 = \frac{p(p-1)(p-2)}{3!}, \cdots$$

$$C_p^r = \frac{p(p-1)\cdots(p-r+1)}{r!}, \qquad (2)$$

这些系数都是整数.以下我们假定 p 为素数.为了把(2)式所给出的这些整数写成整数的形式,需要约去分母 $r!$ 与 $p(p-1)\cdots(p-r+1)$ 中的公因数.但分母不含素因数 p,所以相约后,p 仍然在分子中.这样一来,我们得到定理 1.

定理 1 对整数 x 与 y 及任意素数 p,

$$(x+y)^p \equiv x^p + y^p \pmod{p}. \qquad (3)$$

证 若 p 为素数,则在(1)中的所有的二项式系数,除去第一项和第末项外,其它系数均可被 p 整除.

例 取 $p=5$,我们有

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

中间的项都能被 5 除尽,所以

$$(x+y)^5 \equiv x^5 + y^5 \pmod{5}.$$

从同余式(3)我们可以得到一些重要的推论.首先,让我们把它应用于 $x=y=1$ 的情形,这就给出

$$2^p = (1+1)^p \equiv 1^p + 1^p = 2 \pmod{p}.$$

下一步我们取 $x=2, y=1$,得到

$$3^p = (2+1)^p \equiv 2^p + 1^p,$$

然后,我们利用前面的结果 $2^p \equiv 2 \pmod{p}$,得到

$$2^p + 1^p \equiv 2 + 1 \equiv 3 \pmod{p},$$

所以 $3^p \equiv 3 \pmod{p}$. 再其次,对 $x = 3, y = 1$,我们可得到

$$4^p \equiv 4 \pmod{p}.$$

利用这一方法,就可以依次地证明对所有的值

$$a = 0, 1, \cdots, p-1, \tag{4}$$

有 $a^p \equiv a \pmod{p}$ 成立. 对于特殊情况 $a = 0$ 与 $a = 1$,这个同余式是显然成立的. 因为每一个整数 a 都和式(4)中的某一个数同余 $(\text{mod } p)$,所以我们证明了:

定理 2　对任意的整数 a 和任意的素数 p,有

$$a^p \equiv a \pmod{p}. \tag{5}$$

这一同余定律通常称为费马定理,一些作者为了把它和我们在前面提到的费马大定理或费马猜想相区别而称之为费马小定理.

下面给一个具体例子,说明费马小定理是成立的.

例　对 $p = 13$ 与 $a = 2$,我们可以求得 $13 = 8 + 4 + 1$,所以有 $2^{13} = 2^{8+4+1} = 2^8 \cdot 2^4 \cdot 2^1$. 因为

$$2^4 = 16 \equiv 3 \pmod{13}, 2^8 \equiv 9 \pmod{13},$$

故得　　$2^{13} = 2^8 \cdot 2^4 \cdot 2^1 \equiv 9 \cdot 3 \cdot 2 \equiv 2 \pmod{13}$.

这正是费马同余式所给出的.

根据同余式的消去律,当 a 与模 p 互素时,我们可以在费马同余式(5)的两边消去公因数 a. 这就给出了下面的结果:

定理 2′　若 a 是一个不被素数 p 整除的整数,那么

$$a^{p-1} \equiv 1 \pmod{p}. \tag{6}$$

这一结果也称为费马小定理. 下面给出它的另一证明.

证　考虑序列

$$a, 2a, 3a, \cdots, (p-1)a.$$

这序列有两个性质:1) p 不能整除其中任一数;2) 对模 p,它们彼此不同余. 事实上,

$$ra \equiv sa \pmod{p} \Rightarrow r \equiv s \pmod{p} \Rightarrow s = r.$$

因此,这个序列一定以某种次序模 p 同余于 $1, 2, 3, \cdots, (p-1)$. 所

以，

$$a\cdot 2a\cdot 3a\cdots(p-1)a\equiv 1\cdot 2\cdot 3\cdots(p-1)\pmod p$$

即 $a^{p-1}(p-1)!\equiv(p-1)!\pmod p\Rightarrow a^{p-1}\equiv 1\pmod p$.

例 检查 $5^{38}\equiv 4\pmod{11}$ 是否成立.

解 由费马小定理 $5^{10}=5^{11-1}\equiv 1\pmod{11}$. 从而

$5^{38}\equiv 5^{10\cdot 3+8}=(5^{10})^3(5^2)^4\equiv 1^3\cdot 3^4\equiv 81\equiv 4\pmod{11}$.

作为费马同余式(6)的一个应用，我们回到前面讨论过的毕达哥拉斯三角形上来，证明以下结论：

命题 一个毕达哥拉斯三角形的边长的乘积可被 60 整除.

证 显然，只要对本原三角形来证明就足够了. 根据公式，本原三角形的三边可分别表示为 $x=m^2-n^2, y=2mn, z=m^2+n^2$，这里 m,n 中一个是奇数，一个是偶数. 它们的乘积是

$$P=2mn(m^2-n^2)(m^2+n^2)=2mn(m^4-n^4).$$

易见，当且仅当数 P 被 4,3,及 5 整除时，它才能被 60 整除. 从而我们只需证明 P 分别被 4,3,5 整除. 因为数 m 与 n 中的一个是偶的，所以 $2mn$ 可被 4 整除. 当数 m 与 n 中至少有一个被 3 整除时，P 被 3 整除. 但若 m 和 n 均不能被 3 整除时，P 亦被 3 整除. 这是因为根据费马小定理，从 $(m,3)=1$ 及 $(n,3)=1$ 可推出 $m^2\equiv 1\pmod 3$ 及 $n^2\equiv 1\pmod 3$ 所以

$$m^2-n^2\equiv 1-1\equiv 0\pmod 3.$$

类似地可证明 P 被 5 整除. 若 m 或 n 被 5 整除，那么这是显然的. 如果 m,n 均不被 5 整除，那么再根据费马小定理，我们有

$$m^4-n^4\equiv 1-1\equiv 0\pmod 5.$$

仍然得出 P 被 5 整除的结论. 证毕.

7.3.2 简化剩余系与欧拉函数

这一节我们讨论在完全剩余系中与模互素的整数，由此引入简化剩余系的概念. 为此先引进欧拉函数的概念.

定义　欧拉函数 $\varphi(m)$ 是定义在正整数集合上的函数.它的值等于序列 $0,1,2,\cdots,m-1$ 中与 m 互素的数的个数.

例　$\varphi(6)=2$ 因为在 $0,1,2,3,4,5$,中只有 1 和 5 与 6 互素.$\varphi(7)=6$.因为在 $0,1,2,3,4,5,6$ 中与 7 互素的数是 $1,2,3,4,5,6$.

易见,若 p 是素数,则 $\varphi(p)=p-1$.

定义　如果一个模 m 的剩余类中的数与 m 互素,就把它叫做与 m 互素的剩余类.

例　取 $m=6$,则 $K_5=\{5,11,17,\cdots\}$ 就是一个与 6 互素的剩余类.

在与模 m 互素的全部剩余类中,从每一类中各取一个数组成的集合,叫做模 m 的简化剩余系.

例　序列 1,5 是模 6 的简化剩余系

例　序列 1,2,3,4,5,6 是模 7 的简化剩余系,也是它的完全剩余系.

易见,与模 m 互素的剩余类的个数是 $\varphi(m)$,因此模 m 的每一简化剩余系由与 m 互素的 $\varphi(m)$ 个对模 m 不同余的整数组成.

现在我们来推导一个关于 $\varphi(m)$ 的公式.

设 $m=p_1^{\alpha_1}\cdots p_l^{\alpha_l}$.再设 $M=\{1,2,\cdots,m\}$.在 M 中所有 p_1 的倍数都不与 m 互素,它们是:

$$p_1,2p_1,\cdots,\frac{m}{p_1}\cdot p_1.$$

一共有 m/p_1 个.所以在 M 中与 p_1 互素的数的个数是

$$m-\frac{m}{p_1}=m\left(1-\frac{1}{p_1}\right).$$

把这些数从集合 M 中拿走,剩下的集合叫 M_1,M_1 不再包含 p_1 的倍数.在 M_1 中所有 p_2 的倍数是

$$p_2,2p_2,\cdots,rp_2,\cdots,\frac{m}{p_2}\cdot p_2.$$

这里的任何系数不再被 p_1 除尽,因为 p_1 的任何倍数都不再 M_1 中.在 M_1 中 p_2 的倍数的个数是 $m(1-\frac{1}{p_1})/p_2$.因此在 M 中即与 p_1 互素,又与 p_2 互素的数的个数是

$$m(1-\frac{1}{p_1})-\frac{m}{p_2}(1-\frac{1}{p_1})=m(1-\frac{1}{p_1})(1-\frac{1}{p_2}).$$

这个推导可以继续下去.我们得到定理:

定理 3 若 $m=p_1^{a_1}\cdots p_l^{a_l}=\prod_{i=1}^{l}p_i^{a_i}$ 则

$$\varphi(m)=m(1-\frac{1}{p_1})(1-\frac{1}{p_2})\cdots(1-\frac{1}{p_l})$$

$$=m\prod_{i=1}^{l}(1-\frac{1}{p_i}).$$

证 我们用归纳法来完成定理的证明.$n=1,2$ 的情况已经证明.剩下的只要完成从 k 到 $k+1$ 的过渡就行了.

假定从集合 M 中已经把 $p_1,p_2,\cdots,p_k$ 的倍数都去掉了,剩下的集合用 M_k 表示,它含有的元素的个数是

$$m(1-\frac{1}{p_1})(1-\frac{1}{p_2})\cdots(1-\frac{1}{p_k}).$$

在 M_k 中所有 p_{k+1} 的倍数是

$$p_{k+1},2p_{k+1},\cdots,rp_{k+1},\cdots,\frac{m}{p_{k+1}}\cdot p_{k+1}$$

所有 p_{k-1} 的系数都不被 $p_1,p_2,\cdots,p_k$ 除尽.它们的总个数是

$$\frac{m}{p_{k+1}}(1-\frac{1}{p_1})(1-\frac{1}{p_2})\cdots(1-\frac{1}{p_k})$$

从 M_k 中减去这些数,我们得到

$$m(1-\frac{1}{p_1})\cdots(1-\frac{1}{p_k})-\frac{m}{p_{k+1}}(1-\frac{1}{p_1})\cdots(1-\frac{1}{p_k})$$

$$=m(1-\frac{1}{p_1})\cdots(1-\frac{1}{p_k})(1-\frac{1}{p_{k+1}}).$$

这就是 M 中与 $p_1p_2\cdots p_{k+1}$ 互素的数的个数. 这正是我们要证明的.

特别地,当 p 是素数时

$$\varphi(p^\alpha) = p^\alpha(1 - \frac{1}{p}) = p^{\alpha-1}(p - 1).$$

定理 4 当 $(m, n) = 1$ 时, $\varphi(m \cdot n) = \varphi(m)\varphi(n)$.

证 这是前一定理的直接推论.

事实上,若 $m = \prod_{i=1}^{l} p_i^{\alpha_i}, n = \prod_{i=1}^{k} q_i^{\beta_i}$,则

$$\varphi(m)\varphi(n) = m\prod_{i=1}^{l}(1 - \frac{1}{p_i})n\prod_{i=1}^{k}(1 - \frac{1}{q_i})$$

$$= mn\prod_{i=1}^{l}(1 - \frac{1}{p_i})\prod_{i=1}^{k}(1 - \frac{1}{q_i}) = \varphi(mn).$$

定义 设 $f(x)$ 是定义在一切正整数上的函数. 称 $f(x)$ 是积性函数,如它满足性质:

1) 有一正整数 a,使 $f(a) \neq 0$;

2) 对任意两个互素的正整数 $a, b, f(ab) = f(a)f(b)$.

由定义立刻看出,欧拉函数 $\varphi(m)$ 是积性函数. 积性函数在解析数论中具有十分重要的意义.

7.3.3 欧拉定理

将费马定理推广就得出欧拉定理.

定理 5(欧拉定理) 若 m 是大于 1 的整数, $(a, m) = 1$,则

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

证 设 $n_1, n_2, \cdots, n_{\varphi(m)}$ 是小于等于 m 的简化剩余系,则 $an_1, an_2, \cdots, an_{\varphi(m)}$ 也是模 m 的简化剩余系. 从而

$$an_1 \cdot an_2 \cdot \cdots \cdot an_{\varphi(m)} \equiv n_1 \cdot n_2 \cdot \cdots \cdot n_{\varphi(m)} \pmod{m},$$

或

$$a^{\varphi(m)}\prod n_i \equiv \prod n_i \pmod{m}.$$

但 $(\prod n_i, m) = 1$,因此

$$a^{\varphi(m)} \equiv 1(\bmod\ m).$$

注 在上面的结果中，a 必须与 m 互素.因此推出，同余方程

$$x^{\varphi(m)} \equiv 1(\bmod\ m)$$

恰有 $\varphi(m)$ 个互不同余的解.

7.3.4 对循环小数的应用

先看两个例子.直接作除法可知：

$$\frac{1}{3} = 0.333\cdots;\frac{1}{7} = 0.142857142857\cdots;$$

$$\frac{7}{30} = 0.233\cdots,$$

或记为$\frac{1}{3} = 0.\dot{3};\frac{1}{7} = 0.\dot{1}4285\dot{7};\frac{7}{30} = 0.2\dot{3}$.

前两个小数叫做纯循环小数，后一个小数叫混循环小数.它们的一般定义是：

定义 如果对于无限小数 $0.a_1a_2\cdots a_n\cdots$，其中 a_n 是 $0,1,2,\cdots,9$ 中的一个数，并且从任何一位后不全是 0，能够找到两个整数$s \geqslant 0, t > 0$，使得

$$a_{s+i} = a_{s+kt+i},\qquad i = 1,2,\cdots t, k = 0,1,2,\cdots,$$

我们就称它为循环小数，并简单地把它记作

$$0.a_1a_2\cdots a_s\dot{a}_{s+1}\cdots\dot{a}_{s+t}.$$

对于循环小数而言，具有上述性质的 s 及 t 不只一个.如果找到的 t 是最小的，我们就称 $a_{s+1}a_{s+2}\cdots a_{s+t}$ 为循环节；t 称为循环节的长度；若最小的 $s = 0$，那小数就叫做纯循环小数，否则叫混循环小数.

现在我们给出两个定理，指出什么时候出现纯循环小数，什么时候出现混循环小数.

定理 6 设 $0 < a < b$，$(a,b) = 1$.有理数$\frac{a}{b}$能表成纯循环小数的充分必要条件是$(b,10) = 1$.

证 “$\Rightarrow$”. 若$\frac{a}{b}$能表示成纯循环小数，则

$$\frac{a}{b} = 0.a_1a_2\cdots a_ta_1a_2\cdots a_t\cdots$$

从而

$$10^t\frac{a}{b} = 10^{t-1}a_1 + 10^{t-2}a_2 + \cdots + 10a_{t-1} + a_t + 0.a_1a_2\cdots a_t\cdots$$

$$= q + \frac{a}{b}, q > 0.$$

所以

$$(10^t - 1)\frac{a}{b} = q \Rightarrow bq = a(10^t - 1)$$

由$(a,b) = 1$，可知$b \mid (10^t - 1)$. 因而$(b,10) = 1$.

“$\Leftarrow$”. 设$(b,10) = 1$，今证$\frac{a}{b}$一定可以表示为纯循环小数. 由欧拉定理，一定存在一个正整数t，使得下式成立：

$$10^t \equiv 1(\mathrm{mod}\ b), 0 < t \leqslant \varphi(b).$$

从而
$$10^ta \equiv a(\mathrm{mod}\ b) \Leftrightarrow 10^ta = qb + a,$$

其中 $0 < q < 10^t\frac{a}{b} \leqslant 10^t\frac{b-1}{b} = 10^t(1 - \frac{1}{b}) < 10^t - 1$,

所以
$$10^t\frac{a}{b} = q + \frac{a}{b}. \tag{6}$$

将q表示为下述形式：

$$q = 10^tq_t + 10^{t-1}a_1 + \cdots + 10a_{t-1} + a_t.$$

因为$0 < q < 10^t - 1$，所以$q_t = 0$，且$a_1,a_2,\cdots,a_t$不全为0，因此

$$\frac{q}{10^t} = 0.a_1a_2\cdots a_t.$$

(6) 两边除以10^t，得

$$\frac{a}{b} = \frac{1}{10^t}q + \frac{1}{10^t}\frac{a}{b} = 0.a_1a_2\cdots a_t + \frac{1}{10^t}\cdot\frac{a}{b}$$

反复应用上式,可得

$$\frac{a}{b}=0.a_1a_2\cdots a_ta_1a_2\cdots a_t\cdots=0.\dot{a}_1a_2\cdots\dot{a}_t.$$

例　由$(3,10)=1$,知$\frac{1}{3}=0.\dot{3}$是纯循环小数.同理$\frac{1}{7}=0.\dot{1}4285\dot{7}$也是纯循环小数,但$\frac{7}{30}=0.2\dot{3}$不是纯循环小数,而是混循环小数.因为$(30,10)\neq 1$.

由定理6立即看出当$(b,10)\neq 1$时.有理数$\frac{a}{b}$一定表示为混循环小数.

定理7　设$0<a<b,(a,b)=1,b=2^{\alpha}5^{\beta}b_1,(b_1,10)=1$,$b_1\neq 1$,$\alpha,\beta$不全为零,那么$\frac{a}{b}$可表示为混循环小数,其中不循环的位数是$\mu=\max(\alpha,\beta)$(即$\alpha,\beta$中较大者).

证　可以假定$\mu=\beta\geqslant\alpha$.若$\alpha\geqslant\beta$则证明方法完全一样.用10^{μ}乘$\frac{a}{b}$,得

$$10^{\mu}\cdot\frac{a}{b}=10^{\mu}\cdot\frac{a}{2^{\alpha}5^{\beta}b_1}=2^{\beta}5^{\beta}\cdot\frac{a}{2^{\alpha}5^{\beta}b_1}$$

$$=2^{\beta-\alpha}\frac{a}{b_1}=M+\frac{a_1}{b_1},\tag{7}$$

其中$0<a_1<b_1,0\leqslant M<10^{\mu}$.由

$$2^{\beta-\alpha}a=Mb_1+a_1\text{ 或 }2^{\beta-\alpha}a-Mb_1=a_1,$$

知　$$(a_1,b_1)=(2^{\beta-\alpha}a-Mb_1,b_1)=(2^{\beta-\alpha}a,b_1)=1.$$

因为$(b_1,10)=1,b_1\neq 1$,由定理8,$\frac{a_1}{b_1}$可以展为纯循环小数:

$$\frac{a_1}{b_1}=0.\dot{c}_1\cdots\dot{c}_l.$$

设$M=m_1 10^{\mu-1}+\cdots+m_{\mu}(0\leqslant m_i\leqslant 9)$,(7)式两边除以$10^{\mu}$得

$$\frac{a}{b} = 0.m_1\cdots m_\mu \dot{c}_1\cdots \dot{c}_t.$$

我们还要证明不循环的位数不能小于μ.假定不然,$\frac{a}{b}$又可以表示为

$$\frac{a}{b} = 0.n_1\cdots n_v \dot{d}_1\cdots \dot{d}_s \qquad (v < \mu),$$

则

$$10^v \frac{a}{b} = 10^v n_1 + \cdots + n_v + 0.\dot{d}_1\cdots \dot{d}_s,$$

$$10^v \frac{a}{b} - [10^v n_1 + \cdots + n_v] = 0.\dot{d}_1\cdots \dot{d}_s = \frac{a'_1}{b'_1}.$$

由定理 6 及$(b'_1,10) = 1$,可得

$$10^v \frac{a}{b} = 10^v n_1 + \cdots + n_v + \frac{a'_1}{b'_1}$$

$$= \frac{a'}{b'_1}(\text{化为假分数设分子为 } a').$$

即
$$10^v ab'_1 = a'b,$$
$$5^\mu \mid b \Rightarrow 5^\mu \mid 10^v.$$

所以$v \geqslant \mu$.与假设矛盾,证毕.

习　题

1.计算$\varphi(1001),\varphi(5040),\varphi(36000)$.

2.证明,当$n = 5186$时,$\varphi(n) = \varphi(n+1) = \varphi(n+2)$.

3.证明:1)若n是奇数,则$\varphi(2n) = \varphi(n)$.

　　2)若n是偶数,则$\varphi(2n) = 2\varphi(n)$.

　　3)$\varphi(3n) = 3\varphi(n) \Leftrightarrow 3 \mid n$.

　　4)$\varphi(3n) = 2\varphi(n) \Leftrightarrow 3 \nmid n$.

4. 证明，当 $n=2(2p-1)$ 时，其中 p 和 $(2p-1)$ 都是素数，$\varphi(n)=\varphi(n+2)$.

5. 如果今天是星期一，问从今天起再过 10^{10} 天是星期几？

6. 求 $5^8 \pmod 7$，$5^{12} \pmod{11}$，$1945^8 \pmod 7$，$1945^{12} \pmod{11}$.

§7.4　同余式的应用

7.4.1　在密码学上的应用

同余理论在密码学上有重要的应用. 密码作为军事斗争与政治斗争的一种手段在历史上早就产生了. 信息化社会的到来，使得密码学更加有用. 现在商业信息的往来也需要保密. 通过公共渠道，如电话，电报，电子网络传递信息，希望不被窃取或修改，安全地送到接受者手中，就需要用密文形式传送.

先讲几个名词. 甲方通过公共通道向乙方传输信息，为了防止窃取，甚至篡改，需要将信息改变为秘密形式. 原信息称为明文，明文的秘密形式称为密文. 把明文变为密文的过程叫加密. 知道了密码把密文译为明文的过程叫解密. 密码中的关键信息叫做密钥. 密钥在保密通讯中占有极重要的地位.

一切密码系统都有两部分：1. 一套构成基本密码的通信方法或程序的规则，称为通用系统；2. 一个可变换的密钥，它由数字，单词，词组或句子组成. 在加密时，密钥控制通用系统的步骤，并决定密文的组成. 在解密时，密钥同样地控制着解密的步骤；尽管密码的外部形式和内部构成千差万别，但只有两种基本类形，一种是位移式，一种是代换式. 位移式密码只重新排列或调整明文中字母的顺序，而不改变字母本身. 代换式密码则用其它字母代替明文中的字母而不改变其顺序. 有的密码系统同时使用这两种密码系统.

现在广泛使用各种密码机. 20 世纪 40 年代以来产生出各种电密码机. 许多电密码机都有类似打字机的键盘，并使用一种电转子的装

置产生一系列不同的混合字母,另外一种类似的机器也研制出来了,用它来对众多电码进行加密和解密.

1) 置换密码

下面我们介绍一种简单的,在历史上曾经用过的密码,就是置换密码.我们假定这种密码是用英文发送的.办法很简单,就是把每个字母用另一个字母替换,而形成密文.替换的规则可以是随机的,也可以是系统的.

公元前在高卢战争期间罗马大将凯撒使用的一种密码就是系统置换的密码.置换的规律是:每个字母由它后面第三个字母来替换.例如,

$$A \leftarrow D, B \leftarrow E, C \leftarrow F, D \leftarrow G, \cdots,$$
$$W \leftarrow Z, X \leftarrow A, Y \leftarrow B, Z \leftarrow C.$$

例 Peking University 在这一密码下是

Shnlgj Xglyhuvlwb.

利用同余式的理论,凯撒的密码很容易得到解释.首先把 26 个字母都编上号,按顺序,A 是 01 号,B 是 02 号 …… Y 是 25 号,Z 是 26 号.若用 p 表示明文中的字母编号,而用 s 表示密文中的字母编号,则凯撒密码就可以用同余式写出来:

$$s \equiv p + 3 \pmod{26},$$

式中的 3 就是密钥.解密关键是找到数字 3,找到了 3 之后,立刻就可得到明文.这只需要解同余式

$$p \equiv s - 3 \equiv s + 23 \pmod{26}.$$

更一般些,密码可以由公式

$$s \equiv p + k \pmod{26}$$

给出.为了迷惑企图破译的人,密文通常写为 5 个字母组的形式.于是上面一例可改为

Shnlq jXgly huvlw b

的形式.

对置换密码的解密来说，关键在于确定 k 的值. 解密的方法有两种：一种是穷举法，对 k 一个一个地试，直到出现有明确意义的明文；另一种是根据英文字母出现的频率来进行解密. 在英文中各个字母出现的频率是不同的，如 e 的频率最高，约占 13.04%. 找出字母中出现次数最多的字母，使之对应于 e，然后进行尝试性求解.

1) 仿射变换密码

比置换密码更复杂一些的密码是仿射变换密码，由同余式

$$s \equiv ap + b \pmod{26}$$

给出. 这里 $(a, 26) = 1$，注意到 a 有 12 种可能，b 有 26 种可能，所以仿射变换的总数有 $12 \times 26 = 312$ 种可能. 这比置换密码就复杂一些了. 从而解密也更难一些.

在这种密码中，发送密码的人和接收密码的人使用同一把密钥. 发送者用这把密钥去加密，接收者使用它去解密.

有了计算机这一强大工具之后，上面的密码就变得十分平凡了. 人们不断寻求新的安全密码. 我们来看一看近二十多年来在密码学发生的事情.

3) 现代密码系统

你如何来设计编码(即加密)体系呢？光回答"多加小心"是不够的. 现在的密码分析人员有大量的武器可供使用，既有强大的计算设备，又有复杂的数学和统计技术. 恺撒使用过的那类简单的密码肯定极不安全.

简单的字母替换法被排除之后，还能试用什么办法呢？不管选择哪种方法，同样的危险仍会出现. 只要在你编好的密文中有某种"可被辨认"的模式存在，高级的统计分析方法一般不难破译你的密码. 现在，真正的困难之所在变得明朗了. 为了使信息能安全地到达在接收者那里(可能在数千公里之外)，关键在于，这种隐藏着的规则应埋得足够深，以防被敌人发现.

所有现代的密码体系都要使用计算机. 一般都假定敌人拥有强

大的计算机来分析你的信息，所以你的体系必须足够复杂，以防计算机的攻击. 为了使所设计的密码体系尽可能安全，它们必须由两部分构成：一个加密程序和一把钥匙. 前者是典型的计算机程序，也可能是一台专门设计的计算机. 为了给信息加密，该体系不仅需要这些程序，还要一把选择好的钥匙，它通常是一个秘密选定的数. 加密程序将依赖这把选定的钥匙对信息编码，使得只有知道这把钥匙的人才可能解开所编的密文. 由于安全性依赖于这把钥匙，所以可以有许多人在一段相当长的时期里，使用同一个加密程序. 这就意味着值得花大量的时间和精力来设计这种程序. 不妨看一个有助于理解的类比. 生产保险柜和锁的工厂可以设计一种类型的锁卖给几百个使用者，后者靠自己独特的钥匙来保证安全. 此处所说的"钥匙"可以是用于字码锁的各种字码，于是立即显示出"钥匙"这个词的两种用法间的类似. 正如敌人可能知道你的锁是如何设计的，可是不知道锁的字码仍无法打开你的保险柜. 所以，敌人可能知道你所用的加密体系而无法破译你的编码信息；要想破译就得知道你的钥匙.

在典型的"密钥体系"中，信息发送者和接收者事先要协商好某种密钥，然后利用它互送信息. 只要他们保守钥匙的秘密，该体系应该是安全的. 美国人设计的数据加密标准 DES(Data Encryption Standard) 就属于这种体系. 它的钥匙是个数，其二位进制表示有 56 位. 换言之，这把钥匙是由 56 个 0 和 1 组成的数链. 为什么要这么长的钥匙？好，现在来解释. 实际上没有人对 DES 体系是如何工作的加以保密，一切细节都是公开的. 从理论上讲，任何敌人只要试遍所有可能的钥匙就能找到哪把钥匙在起作用. 就 DES 而言，共有 2^{56} 种可能的钥匙，这个数如此之大，以至想试遍所有的钥匙实际是不可能的. 事实上这个数字还没大到足以提供绝对的安全，不过对任何密码体系都必须兼顾安全性和使用者方便两个因素. 钥匙越长，使用就越不方便.

虽然目前广泛使用着 DES，但该体系有着明显的欠缺. 在使用

前，发送者和接收者必须协商好他们将使用的密钥；因为不愿通过任何通讯渠道传送密钥，他们必须碰面并选定钥匙，或者起码要雇用一位可信任的信使来传送密钥．所以这种体系不适合未曾会面的个人之间的通讯．特别地，它不适合诸如国际间的银行及商务活动，而他们往往需要把保密信息发送给世界各地的从未见过面的人．

1975年，W·迪菲(W.Diffie)和M·赫尔曼(M.E.Hellman)在一篇重要的论文"New Directions in Crytography"(密码术的新方向)中提出一种新型的密码体系：公开密钥的密码系统．其中的编码方法需要两把钥匙而不是一把：一把用于加密，另一把用于解密(就好像一把锁要用一把钥匙把它锁上，用另一把钥匙把它打开)．这种体系的使用方法如下：新的使用者要购买一本供该通讯网络的所有成员使用的标准程序(或专用计算机)．然后他应确定两把钥匙．一把是他的解密钥匙，他应严加保密．另一把钥匙将刊登在网络使用者手册上．为了发送一个信息给网络使用者，需要做的全部工作就是查出那位使用者所公开的那把密钥，用它对信息加密，并发送出去．对任何人而言，知道这把公开的密钥对破译密码是毫无帮助的．解码需要另一把专门解码用的钥匙，而这只有那位接收者知道．所以信息一旦被加了密，连那位信息的发送者也无法破译它！

这种公开密钥的密码学已经成为当今密码学的一个主要研究方向．

看来这办法很不错，但是如何具体地实现这样的体系呢？大家知道，找出大的素数(比如50位数字大小的)相对而言比较容易．而如此大小的两个大素数相乘得出一个和数(大约有100位)也不难．但要把这么大的数分解成两个素因子就难上加难了，无论出于什么意图和目的，事实上后者是不可能做到的．这就是当今普遍使用的公开密钥体系所依赖的基本思想．这一体系是1977年由麻省理工学院的R·里弗斯特(R.Rivest)，A·沙米尔(A.Shamir)和L·阿德勒曼(L.Adleman)设计，现以他们姓氏的首字母命名为RSA体系．需要保密

的用于解密的钥匙(本质上)由两个大的素数组成.使用者要借助于计算机选出它们,而不能在如何公开出版的素数表中去选,后者敌人很容易弄到手!公开的用于加密的钥匙就是这两个素数的乘积.因为不存在快速的因子分解的方法,所以实际上不可能根据公开的加密钥匙重新找到解密钥匙.信息的加密对应于两个大素数相乘(容易),解密对应于相反的因子分解过程(困难).

这就是当今庞大的国际数据通讯网络能安全运行的原因,它依赖的是数学家的一种无能:他们尚未找到大数因子分解的有效方法,同时却能容易地找到大的素数.显然,这类体系的安全性基于因子分解方面的困难,一旦因子分解方面的研究有进展,体系的安全性就受到威胁.

4)新的挑战

这一问题引出了温尼伯啤酒店的一段有趣的对话.1982年秋,在加拿大温尼伯市举行的科学会议期间,两位数学家和一位计算机工程师于某天晚上外出喝啤酒.两位数学家很快转入了如何分解大数的因子的话题,自然碰上了计算问题.计算机工程师听着他们的谈话,指出他设计的一台特殊的计算机可能很容易地克服他们遇到的一个主要的困难.这次的啤酒店碰巧发生的事,对数据安全的研究产生了重大的影响.直到1982年,最好的因子分解方法只能处理约有50位的数字.计算机工程师T·活诺克告诉因子分解专家M·文德利希和G·西蒙斯的信息是,特殊设计的CRAY－1的运算器可以解决他们的一个问题,即能使他们的方法分解有60位至70位的大数.这样,RSA体系的安全就受到了威胁.虽然对付的方法很显然——只要使用每个都有100位的数字的素数,以得到一个有200位数字的公开密钥;但是,这次事先未料到的交谈确实在安全通讯的事业中搅起了一片易变的涟漪.当前的因子分解方面的其它进展也加重了这种“不安全感”.虽说目前一般认为因子分解的上限是有90位的数字,可是大量精巧复杂的数学正瞄准着这个问题,说不定什么时候就

会出现真正的突破.

5) 具体实现

下面谈谈如何来实现 RSA 密码系统.使用者选择一对不同的素数 p, q.它们的乘积 $n = pq$ 叫做加密模.这是一个非常大的数.将它分解为素因数的乘积超出了现代任何计算工具的能力.例如可以选取 p, q 有 100 位数字,这样 n 就有 200 位数字.n 确定以后,再随机地选取一个正数 k,叫做加密指数,使它满足 $(k, \varphi(n)) = 1$.然后把这一对数刊登在网络使用者手册上,作为加密的钥匙用.这样,使用公共网络的任何人都可以用它作为加密的钥匙,将信息加密,并发送给那位网络使用者.注意,n, k 是公开的,而 n 的因数 p, q 却是保密的.加密指数 k 的选择有多种可能.一种方便的办法是取 $\varphi(n) + 1$ 的任何素因数.

加密的过程是这样的.首先把信息转变为一个整数 M.常用的办法是,令

A = 01	I = 09	Q = 17	Y = 25	3 = 33
B = 02	J = 10	R = 18	Z = 26	4 = 34
C = 03	K = 11	S = 19	' = 27	5 = 35
D = 04	L = 12	T = 20	. = 28	6 = 36
E = 05	M = 13	U = 21	? = 29	7 = 37
F = 06	N = 14	V = 22	0 = 30	8 = 38
G = 07	O = 15	W = 23	1 = 31	9 = 39
H = 08	P = 16	X = 24	2 = 32	! = 40.

用 00 表示字之间的空位.按照这种办法,信息

The brown fox is quick

就译为数串

$$M = 2008050002181523140006152400091900172109031128.$$

我们假定明文产生的数 $M < n$.要是原文太长,不能用一个小于 n 的 M 表示它,就把 M 分成适当大小的几段,使每一段都小于 n,然后

对每一段分别加密.

在网络手册上找出接收者的密钥(n,k).将明文数 M 变为密文数 r.r 是这样得到的:求出 M 的 k 次幂,然后用模 n 去简化,即

$$M^k \equiv r(\bmod n).$$

在高速电子计算机上,加密一个有 200 位数字的信息只要几秒钟就能完成.

把密文数 r 发送出去.

现在看网络的另一端.密文接收者首先要做的事是,求出密文复原指数 j.j 满足方程

$$kj \equiv 1(\bmod \varphi(n)).$$

因为$(k,\varphi(n))=1$,所以同余方程有唯一解.这个解很容易得到.由欧拉定理

$$k^{\varphi(\varphi(n))} \equiv 1(\bmod \varphi(n))$$

与前一方程结合起来,得到

$$k^{\varphi(\varphi(n))} \equiv kj(\bmod \varphi(n)) \Rightarrow j \equiv k^{\varphi(\varphi(n))-1}(\bmod \varphi(n)).$$

这个复原指数 j 就求出来了.但是只有知道 $\varphi(n)=(p-1)(q-1)$ 的人才能通过 k 找到这个复原指数,也就是只有知道 p,q 的人才能找到它.只知道公开的密钥是无能为力的.

现在只需通过 r,计算 $r^j(\bmod n)$ 就能重新找到 M.因为 $kj=1+\varphi(n)t$,其中 t 是某个整数所以

$$\begin{aligned} r^j &\equiv (M^k)^j \equiv M^{kj} \equiv M^{1+\varphi(n)t} \\ &\equiv M(M^{\varphi(n)})^t \equiv M\cdot 1^t \equiv M(\bmod m), \end{aligned} \tag{1}$$

这里我们用了$(M,n)=1$ 这个条件.这样一来,我们找到了明文数 M.

假定$(M,n)=1$是为了使用欧拉定理.只有当 $M=p$,或 $M=q$ 时才会出现M 与 n 不互素的情况.出现这种情况的概率太小了,在实际上是不会发生这种情况的.

这种密码的巧妙之处在于,加密时只用到 n,只有解密的人才用

到 p,q.

例 为了使读者对RSA公开密钥的系统有所熟悉,这里给出一个较为详尽的例子.首先选取两个素数 $p = 29, q = 53$.于是加密模是 $n = 29 \cdot 53 = 1537$.而

$$\varphi(n) + 1 = 52 \cdot 28 + 1 = 1457 = 31 \cdot 47.$$

将加密指数取为 $k = 47$.因为复原指数 j 满足同余式 $kj \equiv 1(\bmod \varphi(n))$,所以 $j = 31$.假定明文信息是

No way,

按照前面指出的将它变为明文数:

$$M = 141500230125.$$

我们还要求明文数段不超过1537.在这一限制下,将 M 分为四段,每段的数字都是三位的.它们分别是

$$141, \quad 500, \quad 230, \quad 125.$$

然后分别对它们加密.第一段是141,它的密文是

$$141^{47} \equiv 658(\bmod 1537),$$

把其它三段密文也找出来,全部密文是0658 1408 1250 1252.

现在看网络的另一边.接收者已经知道复原指数 $j = 31$.通过计算可以把明文数找出来:

$$658^{31} \equiv 141(\bmod 1537).$$

其它三段也可以类似地找出来.这样我们就解出了原文.

对RSA系统而言,解密的直接途径就是分解 n,一旦找到了 n 的素因数,从

$$\varphi(n) = (p-1)(q-1)$$

就可以求出复原指数 k,明文也就立刻得到了.分解一个合数比判断一个数是合数还是素数要困难得多.使用目前最快的计算机,判断一个200位的数是不是素数只要十分钟就够了.但是要分解一个200位的合数要花费巨大的劳动,时间之长是难以想象的.假定一微秒(10^{-6} 秒)做一次运算,分解一个200位的合数要 3.8×10^{9} 年.所以

RSA 系统是相当安全的.

7.4.2 素数鉴别

虽然有关素数的大部分经典问题还没有解决,但检验一个数是否素数的方法在最近几年有了巨大进步.你或许会说,检验素数有什么难?确实,看一个数是不是素数,有一种非常自然而直接的方法,这就是我们常用的试除法.给你一个数,譬如说 n,你先看 2 是否能整它,若能,则 n 不是素数,任务完成;若不能,你再用 3 试除 n,若 3 能整除 n,则 n 不是素数,事情了结;若不能,再用 5 去除 n.依次类推.如果试到$\sqrt{n}$ 还没有找到能整除 n 的数,那么 n 一定是素数.

这一方法对检验不太大的数是挺实用的.但若数字太大,它就变得十分笨拙.假设你在一个快速计算机上使用高效的程序进行试除.对于一个 10 位数字的数,运行程序几乎瞬间就能完成.对一个 20 位的数就麻烦一点了,需要两个小时.对于一个 50 位的数,则需要 100 亿年.这已经大得不可想象.当然,这不是对非常大的数作无关紧要的计算.前面说过,介于 60 位到 100 位之间的数是今日密码体系中最安全的一种密码所需要的.

如何确定一个 100 位的数是否素数呢?目前可用的最好的方法是 1980 年左右得到的.数学家阿德勒曼,鲁梅利,科恩和伦斯特拉研究出一种非常复杂的方法.现在以他们的名字的第一个字母命名为 ARCL 检验法.在上面提到的那类计算机上进行 ARCL 检验,对 20 位的数只需 10 秒种,对 50 位的数用 15 秒,100 位的数用 40 秒.如果要检查一个 1000 位的数,一个星期也就够了.这种检验依靠相当多的高深的数学,超出了普通的大学数学的范围.所以这里不能给以完全的回答.但解释该方法的中心思想倒不难,它涉及到费马小定理.

1) 判别法

用费马小定理来判断一个大数是不是素数是一个非常有效的方法.费马定理说,若 n 是素数,则

$$a^n \equiv a \pmod{n}.$$

这样一来,如果对一个给定的 n,有一个 a,使(1)不成立,则 n 一定不是素数.换言之,n 一定是一个合数.

例 判断 $n = 117$ 是不是素数.

解 为了使用费马小定理,我们取 $a = 2$ 看

$$2^{117} \equiv 2 \pmod{117}$$

是否成立,直接计算知道

$$2^7 = 128 \equiv 11 \pmod{117},$$

而 $117 = 7 \times 16 + 5$,所以

$$2^{117} = 2^{7\cdot 16+5} = (2^7)^{16} \cdot 2^5$$
$$\equiv (11)^{16} \cdot 2^5 \equiv (11^2)^8 \cdot 2^5 \equiv (121)^8 \cdot 2^5 (\mathit{mod}\ 117),$$
$$2^{21} = (2^7)^3 \equiv (11)^3 \equiv 121 \cdot 11 \equiv 4 \cdot 11 \equiv 44 \pmod{117},$$

这就指出 $$2^{117} \equiv 44 \not\equiv 2 \pmod{117}.$$

因此 117 不是素数,而是合数,事实上 $117 = 13 \cdot 9$.

2)费马小定理的逆定理

上面的问题又使我们提出这样的问题:费马小定理的逆定理成立不成立?也就是说,如果

$$a^n \equiv a \pmod{n},$$

能不能判断 n 是素数?下面给出一个例子说明费马小定理的逆定理不成立.我们需要一个引理.

引理 若 p, q 是两个不同的素数使得 $a^p \equiv a \pmod{q}$,$a^q \equiv a \pmod{p}$,则 $a^{pq} \equiv a \pmod{pq}$.

证 根据假设,

$$a^q \equiv a \pmod{p} \Rightarrow a^{pq} \equiv a^p \equiv a \pmod{p}$$
$$\Rightarrow p \mid a^{pq} - a.$$

同理,$p \mid a^{pq} - a \Rightarrow a^{pq} \equiv a \pmod{pq}$.

费马小定理的反例 $2^{340} \equiv 1 \pmod{341}$,$341 = 11 \cdot 31$.

解 注意到 $2^{10} = 1024 = 31 \times 33 + 1$，于是

$$2^{11} = 2 \cdot 2^{10} \equiv 2 \cdot 1 \equiv 2 \pmod{31}.$$

$$2^{31} = 2(2^{10})^3 \equiv 2 \cdot 1^3 \equiv 2 \pmod{11}.$$

由引理 $2^{11 \cdot 31} \equiv 2 \pmod{11 \cdot 31} \Leftrightarrow 2^{341} \equiv 2 \pmod{341}$，

或

$$2^{340} \equiv 1 \pmod{341}.$$

这个结果说明，费马小定理的逆是不正确的.

有趣的是，中国数学家早在2500年前就猜测一个数 n 是素数的充要条件是

$$n \mid (2^n - 2).$$

这一命题在 $n \leqslant 340$ 时是正确的.上面的例子说明这个猜测是错误的.数341是1819年才发现的.由此引出了下面的定义.

定义 一个合数 n 叫做伪素数，如果 $n \mid (2^n - 2)$.

可以证明伪素数有无限多个.前4个伪素数是341，561，645和1105.伪素数虽然无穷多，但出现的频率比素数小得多.在1000以内只有3个，在100万之内只有245个，而100万内的素数有78492个.第一个偶的伪素数是

$$161038 = 2 \cdot 73 \cdot 1103,$$

是1950年发现的.

当你用费马小定理去检验素数时，发现 n 能整除 $(2^n - 2)$，你能得到的结论是：n 是素数，或者是伪素数.这时素数的可能性大，因为相对于素数而言，伪素数少得多.

还需指出，当你用费马小定理检查素数时，用其它数，比如3和5代替2，仍会出现伪素数.这种使你在检验素性的问题时得不到绝对肯定的解答.

存在合数 n，对一切整数 a，具有性质：

$$a^n = a \pmod{n},$$

这种 n 中最小的是561.这种例外的数叫做绝对伪素数.

ARCL 检验改进了费马检验，它不再受伪素数的愚弄，这一改进

需要许多高深的数学.

7.4.3 星期数

现在我们利用同余式的理论来计算某一个特定的日子是星期几.例如,香港回归日 1997 年 7 月 1 日是星期几,或者中华人民共和国成立日 1949 年 10 月 1 月是星期几.如果一年的天数可被 7 整除,那么所有的日期在每一年中总有相同的星期数.这样日历的编制将大大简化.未来公历改革后可能会出现这一可喜情况,但是目前却不行.一年的天数是 $365 \equiv 1 \pmod 7$,而闰年的天数是 $366 \equiv 2 \pmod 7$,这表明,在平年一个给定日期的星期数在下一年要加 1.碰到闰年这一规律就被破坏了.

现在我们给出一个方便的计算公式,为此规定:

星期日 = 0,星期一 = 1,星期二 = 2,星期三 = 3,

星期四 = 4,星期五 = 5,星期六 = 6.

另外注意到闰年所加的一天不在一年的开头,也不在一年的末尾,而是在 2 月的 29 日.所以为方便计,我们把 3 月当作第一个月;4 月算作第二个月,把 1 月算作上一年的第十一个月,2 月算作上一年的第十二个月.这一约定对计算星期数是有好处的.作了这样的规定之后 1997 年 7 月 1 日就要写为 1997 年"5 月"1 日,而 1998 年 1 月 1 日就要写为 1997 年"11 月"1 日了.

现在假定我们有一个给定的日期:第 N 年 m 月 d 日.年份数 $N = c \cdot 100 + y$,c 是世纪数,y 是在这一世纪中的年代数,月份数 m 按上面的规定计算.用 W 表示这一日期的星期数.W 的计算公式如下:

$$W \equiv d + \left[\frac{1}{5}(13m - 1)\right] + y + \left[\frac{1}{4}y\right] + \left[\frac{1}{4}c\right] - 2c \pmod 7,$$

公式中的方括号表示不超过这个数的最大整数.

例 香港回归日;1997 年 7 月 1 日.这里

$$c = 19, y = 97, m = 5, d = 1,$$

代入公式得

$$W \equiv 1 + \left[\frac{64}{5}\right] + 97 + \left[\frac{97}{4}\right] + \left[\frac{19}{4}\right] - 2 \times 19$$

$$\equiv 1 + 12 + 97 + 24 + 4 - 38 \equiv 100 \equiv 2 \pmod 7.$$

香港回归日是星期二.

例　中华人民共和国成立日:1949 年 10 月 1 日.这里

$$c = 19, y = 49, m = 8, d = 1,$$

代入公式得

$$W \equiv 1 + \left[\frac{103}{5}\right] + 49 + \left[\frac{49}{4}\right] + \left[\frac{19}{4}\right] - 2 \times 19$$

$$\equiv 1 + 20 + 49 + 12 + 4 - 38 \equiv 6 \pmod 7.$$

中华人民共和国成立日是星期六

这里值得注意的是,目前国际通用的公历是教皇格里高利十三实行的.当时他召集了很多学者和僧侣讨论历法改革问题,决定采用业余天文学家利里奥的方案,每四百年去掉三个闰日.公元 1582 年格里高利颁发了改历的命令:

1) 把 1582 年 10 月 4 日以后的一天改为 1582 年 10 月 25 日;

2) 那些世纪数不能被 4 整除的世纪年(如 1700、1800,1900,2100,…)不再算作闰年,仍算作平年.

这两条规定至为重要.第一条规定实质上把春分日固定在 3 月 21 日左右,解决了日历与天时不合的矛盾.第二条规定把历法的精度大大地提高了一步,保证这种历法在相当长的时期内也能适用.根据这项规定,400 年中共有 97 个闰年,总日数为

$$365 \times 400 + 97 = 146097(\text{天}).$$

因此平均每年的长度为

$$146097 \div 400 = 365.2425(\text{天}).$$

这与回归年实测值 365.2422 天相差只有 0.0003 天.换句话说,要经过三千三百多年,这两者才有一天的相差.格里高利历显然要精密的多了.正是由于格里高利历精度很高,因此,先是欧州后是世界各国

陆续地采用了这个历，这就是现在所通称的"公历". 我国采用公历是在辛亥革命后的 1912 年.

公历的回归年长度取为 365.2425 天，我国早在公元 1199 年使用的南宋"统天历"中就采用了这个数值，比公历早了三百八十多年.

7.4.4　公式的证明

证明分为两步，1) 先求出第 N 年 3 月 1 日的星期数；2) 再求出第 N 年 m 月 d 日的星期数.

1) 任取一年，比如说以 1600 年作为开始的一年. 并把这一年的 3 月 1 日的星期数记为 d_{1600}.

若没有闰年，那么用每过一年在 d_{1600} 上加 1 的办法就可得到第 N 年 3 月 1 日的星期数，因而相应的数为

$$d_{1600} + (100c + y - 1600)(\mathrm{mod}\ 7).$$

考虑到闰年，并假定每四年闰一次，那么应该在这个数上在加上

$$\left[\frac{1}{4}(100c + y - 1600)\right] = 25c - 400 + \left[\frac{1}{4}y\right].$$

这就多了一点，因为世纪年通常不是闰年，所以还应当在这个量中减去 $c - 16$. 但当世纪数 c 可被 4 整除时，第 $100c$ 年仍是闰年，所以我们还需要加上最后一个校正项

$$\left[\frac{1}{4}(c - 16)\right] = \left[\frac{1}{4}c\right] - 4.$$

把这几个式子并在一起，就得出第 N 年 3 月 1 日的星期数

$$d_N \equiv d_{1600} + 124c + y - 1988 + \left[\frac{1}{4}c\right] + \left[\frac{1}{4}y\right] (\mathrm{mod}\ 7).$$

再按模 7 简后化，得到

$$d_N \equiv d_{1600} - 2c + y + \left[\frac{1}{4}c\right] + \left[\frac{1}{4}Y\right] (\mathrm{mod} 7). \qquad (2)$$

把这个公式应用于 1997 年. 这一年的 3 月 1 日是星期六. 因此

$d_{1997}=6$. 这时

$$c=19,\left[\frac{1}{4}c\right]=4,y=97,\left[\frac{1}{4}y\right]=24.$$

我们得到

$$d_{1997}=6\equiv d_{1600}-2\times19+97+4+24\equiv d_{1600}+3(\mathrm{mod}7),$$

所以

$$d_{1600}=3.$$

1600 年 3 月 1 日是星期三. 这一结果代入公式(2),我们得到任意一年的 3 月 1 日的星期数的公式

$$d_N=3-2c+y+\left[\frac{1}{4}c\right]+\left[\frac{1}{4}y\right] \tag{3}$$

2) 我们来确定从 3 月 1 日到该年任意其它一天的天数(mod7). 因为各月天数不同,还需要费点周折. 先求其它月份第一天的星期数.

3 月有 31 天,所以 4 月 1 日的星期数应加 3. 4 月有 30 天,所以 5 月 1 日的星期数必须加 3 + 2 = 5. 这样继续下去,就可得到下面的加法表

月份	所加天数	月份	所加天数
3	0	9	16
4	3	10	18
5	5	11	21
6	8	12	23
7	10	1	26
8	13	2	29

表中数虽然不规则,但每月的平均增长是$\frac{29}{11}=2.6\cdots$ 因为第一项是 0,所以 m 个月的增长数应是 $m\times2.6$ 减去 2.6,然后取其整数部分,即$[2.6m-2.6]$,但这结果不完全正确. 通过修正被减项,我们得到了正确的表达式

$$[2.6m-2.2]=\left[\frac{1}{5}(13m-11)\right],m=1,2,\cdots,12. \tag{4}$$

在上式中核算 $m = 1,2,\cdots,12$ 各值，你会发现它们正是表中的数值.这样我们就得到了第 m 月第一天的星期数，就是把(4) 加到(3)上去.这个月第 d 日的星期数应加上 $d-1$.调整一下各顺序，我们就得到所求公式.

7.4.5 循环赛程排列

现在我们用同余理论来安排各种循环比赛的程序表.这是一个很有实用价值的问题.

假定有 N 队或 N 个选手参加比赛.当 N 为奇数时总有一队要轮空.我们可以用这样的办法来克服这一困难;加进一个假想的队 T_0.然后安排包括 T_0 在内的 $N+1$ 个队的比赛程序表.在每轮比赛中，安排和 T_0 比赛的队轮空.所以我们总可以假定有偶数个队参加比赛，以下假定 N 是偶数，并给每队一个编号 $x = 1,2,\cdots,N$.每队比赛的总场数是 $N-1$，一共有

$$S = (N-1) + (N-2) + \cdots + 1 = \frac{N(N-1)}{2}$$

场比赛.

现在假定 x 属于集合

$$\{1,2,\cdots,N-1\}, \tag{5}$$

在第 r 轮比赛中，以 x_r 表示与 x 队进行比赛的队的编号，x_r 属于集合(5).我们用同余式

$$x + x_r \equiv r(\mathrm{mod}\ N-1) \tag{6}$$

来确定 x_r，就是使 x 与它的对手 x_r 之和在第 r 轮比赛中同余于 r. r 不同，x 的对手也不同

例如，有六个队参加比赛.那么 $N = 6$，x 属于集合 1,2,3,4,5.在第一场比赛中，x 的对手是 x_1，x 与 x_1 满足同余式

$$x + x_1 \equiv 1(\mathrm{mod}\ 5).$$

当 x 是第一队时，$x_1 = 5$，即它的对手是第五队.当 x 是第二队时，$x_1 = 4$，即它的对手是第四队.

在第二场比赛中，x 的对手是 x_2，x 与 x_2 满足同余式

$$x + x_2 \equiv 2 \pmod 5.$$

当 x 是第四队时，$x_1 = 3$，即它的对手是第三队. 当 x 是第五队时，$x_1 = 2$，即它的对手是第二队.

这样安排就会使不同的队有不同的对手，事实上，由

$$x + x_r \equiv r \equiv x' + x_r \pmod{N-1} \Rightarrow x \equiv x' \pmod{N-1},$$

因为 x, x' 属于同一集合(5)，所以 $x = x'$.

唯一的产生麻烦的地方是 $x = x_r$ 的情况. 由(6)，必有

$$2x \equiv r \pmod{N-1}, \tag{7}$$

不难看出在(6)中只有一个 x 出现这种情况. (7)在(5)中只有一个解，即

$$x = \frac{r}{2}, \qquad r\ 为偶数.$$

$$x = \frac{r + N - 1}{2}, \qquad r\ 为奇数.$$

这个例外队叫做 x_0，让它与第 N 队比赛. 这样一来，对集合(5)中的每一队，我们都指定了它在第 r 轮比赛的对象.

习　题

1. 用凯撒的加密方法对信息 RETURN HOME 加密.

2. 设凯撒密码的密文是 KDSSB ELUWKGDB，求明文.

3. 用仿射变换 $C = aP + b \pmod{26}$ 对信息 NUMBER THEORY IS EASY 进行加密.

4. 当 RSA 系统的密钥是 $(n, k) = (3233, 37)$ 时，求复原指数.

5. 以 $(n, k) = (2419, 3)$ 作密钥，用 RSA 加密系统对信息 GOLD MEDAL 进行加密.

第八章　分形与混沌

明天不熟悉分形的人，将不能认为是科学上的文化人.

J. Wheeler

数学的伟大使命是在混沌中发现有序.

N. Weiner

混沌和分形…是数学兄弟. 它们都与不规则结构斗得难分难解.

I. Stewart

一切都是有秩序中的无秩序.

罗曼．罗兰

§8.1　漫游分形

8.1.1　引言

从古希腊以来，人们研究直线，圆，椭圆，双曲线等规则图形. 这是欧氏几何，解析几何及微积分所研究的主要图形. 30 年前诞生了一门新的几何学，称为分形几何. 它研究的却是自然界中常见的，不规则的，不稳定的，变化莫测的现象. 分形几何的创始人芒德布罗(B. B. Mandelbrot)在他的名著"自然界的分形几何"一书中说：

云彩不是球，山岳不是锥体，海岸线不是圆，树皮不是光滑的，闪电也不是沿直线传播的.

他还指出，自然界的许多物体的形状及现象的复杂性是寻常的事，但欧氏几何却把它们抛在一边，不加理会. 分形几何为阐述这类复杂性提供了全新的概念和方法，30 年来取得了惊人的成就，而成

为当代最具有吸引力的科学研究领域之一.

分形是 fractal 的译名,这个词是蒙德尔布罗根据拉丁词 fractus 的词首与英文 fractional 的词尾合成的一个新词,用以描述那种不规则的,破碎的,琐屑的几何特征;既可当名词用,又可当形容词用.分形概念并非纯数学抽象的产物,而是对普遍存在的复杂几何形态的科学概括,有极为广泛的实际背景.自然界中分形体无处不在,起伏蜿蜒的山脉,坑坑洼洼的地面,曲曲折折的海岸线,层层分叉的树枝,支流纵横的水系,变幻莫测的浮云,地质学中的复杂褶皱,遍布动物周身的血管等等,都是自然界中的分形.就是社会历史领域也不乏分形现象,只是不那么直观罢了.

分形是相对于整形而言的.传统几何学描述的对象是由直线或曲线,平面或曲面,平直体或曲体构成的各种几何形状,称为整形.整形的基本特征是具有光滑性,即可微性,至少是分段或分片光滑的,除少数例外点或点集,形体都是可微(可用可微的函数来描述)的.分形的基本特征是不可微的,不光滑的,甚至是不连续的.传统观点把自然界想象成各种规则形体的总和,但普遍存在的几何对象大多数是分形,整形倒是一种例外.

几何学讲的整形是严格定义的数学对象.分形也应当建立严格的数学定义.但目前尚无可以普遍接受的严格定义.非正式地讲,一种几何图形,如果它的组成部分与整个图形有某种方式的相似性,就是分形.

对"分形"的定义可以用和生物学中对"生命"定义的同样方法处理.在生物学中"生命"并没有严格和明确的定义,但可以列出一系列生命物的特性:象繁殖能力,运动能力以及对周围环境的相对独立的存在能力等.大部分的生物都具有上述的特性,虽然一些生物对上述某些性质有例外.同样,对分形似乎最好把它看成具有下面列出的性质的集合,而不去寻找精确的定义.因为这种定义肯定几乎总要排除掉一些有趣的情形.

称一个集合 F 是分形，如果它具有下面的典型性质：

1）F 具有精细的结构，即有任意小比例的细节；

2）F 非常不规则，它的整体和局部都不能用传统的几何语言来描述；

3）F 通常具有某种自相似的性质，这种自相似的性质可能是近似的或是统计的；

4）一般说来，F 的分形维数大于它的拓扑维数；

5）在大多数令人感兴趣的情形，F 以非常简单的方法定义，并且可能由迭代产生.

下面我们考察一些具体的分形.

8.1.2 海岸线的长度

为了对分形的概念有所了解.我们举几个具体的例子.首先看一看海岸线长度的测量问题.

英国有一位叫里查逊(L. F. Richardson)的科学家，为了研究海岸线查阅了西班牙，葡萄牙，比利时，荷兰的百科全书.之后，他惊奇地发现，各国各自测量的共同的国境河岸长度竟相差 20%.真是见鬼了！于是他向世界提出了海岸线的问题.

1967 年蒙德尔布罗在一篇叫"英国海岸线有多长，统计自相似性与分数维"的文章中对海岸线长度的问题做了分析.他指出：

"事实上任何海岸线在某种意义上都是无穷地长，从另一种意义说，答案取决于你所用的尺的长度.如果用 1 公理的尺子沿海岸测量，小于 1 公理的那些弯弯曲曲就会被忽略掉.若用 1 米的尺子，会得出较长的海岸线，因为它会捕捉到一些曲折的细部.反之，若用一种在卫星上观察的方法，一定会得出较短的海岸线长度.再反过来，从蜗牛爬过每一个石子来看，这岸线必然长得吓人.

或许有许多人会认为不断增加的岸线长度最后会收敛于一个特定的最后数值，即海岸线的真正长度.可是，假如海岸线是一种欧几里得图形，例如圆，直线，那是可能的.由小线段不断地取更小的段可

以真正地收敛于圆周或线段的长度.事实上,随着测量尺度的变小,测出的海岸线长度无限增大.小湾内有小湾,小半岛之外有小小半岛,直到原子的尺寸方才达到终点,而那里的尺度是无限地复杂."

8.1.3 科克曲线

现在考虑一种更简单的模型,称为科克曲线,或雪花曲线(图 8－1).它是 1904 年瑞典科学家科克(Helge von koch)所描述的:

"一正三角形,每一边的长度是 1,现在在每个边的正中间 1/3 处再凸出造一个正三角形,小三角形在三个边上出现,使原三角形变成六角形;再在六边形的 12 条边上重复进行中间 1/3 处凸出一正三角形的过程,得到了 $4\times12=48$ 边形;每边的正中间还可以再在 1/3 处凸出一小正三角形,如此至于无穷.其外缘的构造越来越精细.它好象是一片理想的雪花."

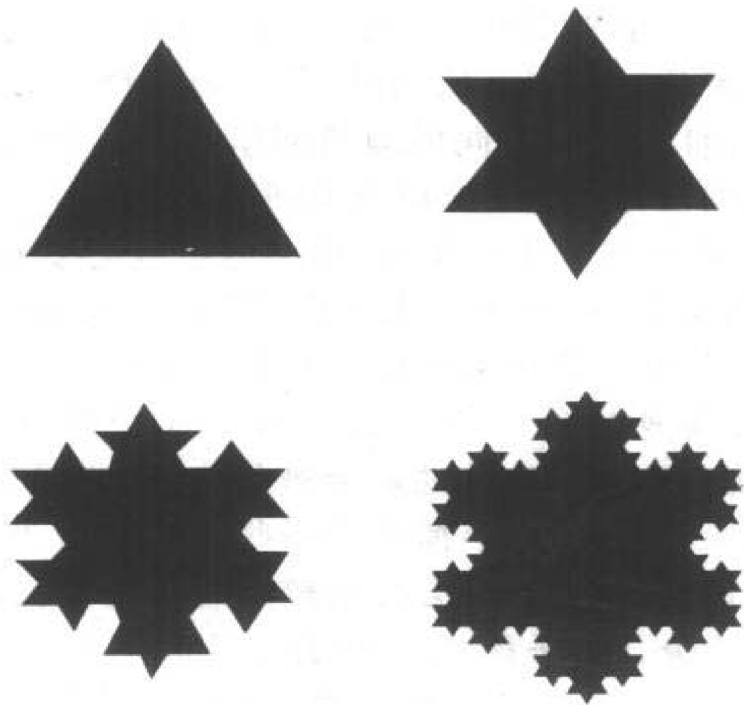


图 8－1　科克雪片的头四个阶段的构造

科克曲线有一些有趣的特性.它是一条连续的闭合曲线,自身不相交.它的总面积是有限的,永远小于原正三角形的外接圆的面积.但它的长度却是无穷地长,因为每次变换后的长度都是原来长度的 4/3 倍,所以当这个过程无限进行下去时,边缘的长度为$3\times(4/3)\times(4/3)\times(4/3)\times\cdots\rightarrow\infty$.这是一个似乎自相矛盾的结果,在有限的空间中有无穷长的线,这的确令人惊讶.

这条曲线有一个特点,那就是局部与整体的相似性.这就是曼氏的分形几何中的一个极重要的概念:自相似性.即取分形图形的任一部分进行适当放大,便仍可得到与原来整个图形相似的图形.现代"混沌"理论的研究发现,"混沌"具有外表混乱而实际上无穷自相似的嵌套结构.这样,"分形"与"混沌"的研究便在"自相似性"这一点上汇合在一起了.

8.1.4 皮亚诺曲线

经典的几何方法和计算方法已经不适合用来研究分形,需要寻找另外的方法.研究分形几何的主要工具是它的许多形式的维数.下面给出一种反映比例性质和自相似性质的分形维数计算法.

在谈分数维以前,首先谈谈什么是维数.

我们根据经验得知,点是 0 维的,直线是 1 维的,平面是 2 维的,而我们居住的空间却是 3 维的.如果像相对论那样,把时间和空间作为同等处理,那么我们居住的空间就是 4 维的了.所有这些经验的维数都是整数,其数字与单独挑选的变数数和自由度的数目是一致的.也就是说,直线上的任意点可用一个实数表示,平面上的任意点可用 2 个实数组表示.如果把维数作为自由度的数目,那么对任意非负的整数 n,在数学上可以考虑 n 维空间.实际上,在处理质点系运动时,把坐标和运动量看作独立变数,把 n 个粒子系看作 $6n$ 维空间中一个点的运动是力学的基础.

把直线弯一下,就得到一条曲线.可见,曲线也是一维的.把曲面弯一下,就得到一张曲面.可见,曲面也是二维的.

把自由度作为维数的设想是非常自然的，而且也没有特别使人产生疑问的地方．大多数人都是这么理解的，对数学的大部分分支这种理解也是足够的，并没有遇到什么问题．但是，当人们深入地研究曲线和曲面定义的时候遇到了问题．早在100年前（1890年），对经验维数已提出了较深刻的疑问．这是意大利数学家皮亚诺发现的．他构造了一种曲线，现在叫做皮亚诺曲线，皮亚诺曲线对经典维数提出了挑战．

皮亚诺（Peano Guiseppe，1858.8 – 1932.4）是意大利数学家、符号逻辑的奠基人，毕生致力于建立数学基础和发展形式逻辑语言．他的著名的工作有：自然数的公理系统，现在称为皮亚诺公理系统；给出了曲线和曲面的定义，他构造了皮亚诺曲线；他还是“国际语”的创始人．

皮亚诺曲线既对经典维数提出挑战，又具有分形结构，值得讲一讲它的构造．皮亚诺曲线可定义为图8 – 2中折线的极限．从图上

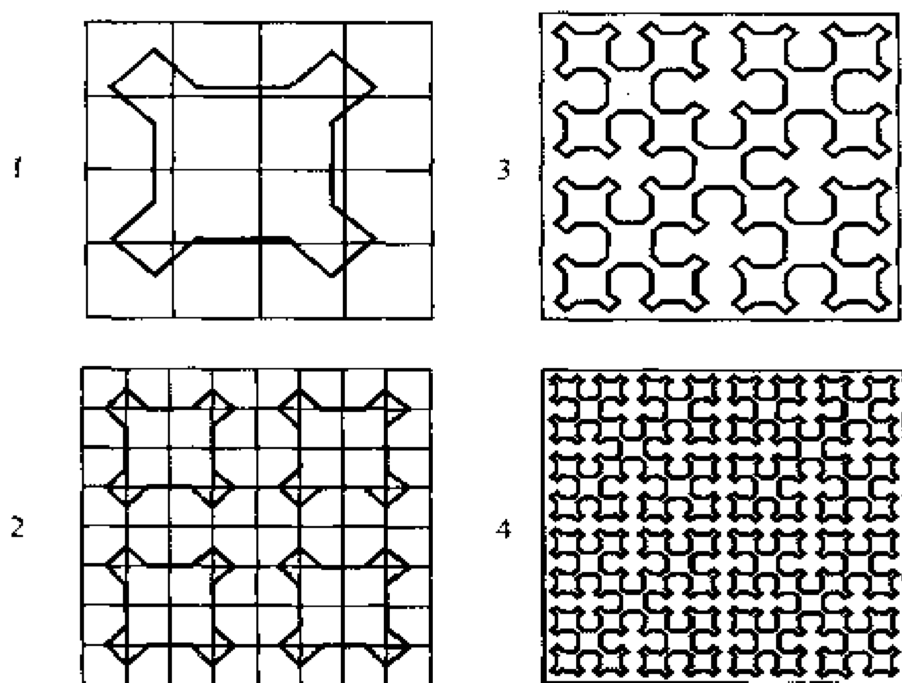


图8 – 2　线的头四级

也可看出,这条曲线完全把一个正方形给覆盖了.

这真是一条怪曲线.怪在一条一维曲线竟把二维区域给盖住了.这样,用一个实数就可以表示二维区域的点,平面不就变成一个自由度了吗?皮亚诺曲线的考虑方法可适用于3维以上,同时也可用一个实数表示 n 维空间中的任意点.也就是说,如果从自由度角度考虑,也可把 n 维空间看作1维空间.这显然就对经典维数提出了挑战.

这条曲线具有自相似性,并且处处都不可微,它是分形的一个典型例子.

8.1.5 分数维

为了避免这一矛盾,必须从根本上重新考虑维的意义.现已提出不少有关维的定义,其中最易理解的是被叫作相似性维数的量.

根据相似性,现在来看看线段,正方形和立方体的维数.首先,如图8-3把各图形的边分成二等分,当然,线段是一半长度的二个线段之并.正方形是边为原来边长 $\frac{1}{2}$ 的4个正方形之并,而立方体则为8个小立方体之并.也就是说,线段,正方形,立方体可被看成为分别由2,4,8个把全体分成的相似形组成.2,4,8数字可改写为 $2^1, 2^2, 2^3$,这里出现的指数1,2,3则分别与其图形的经验维数相一致.

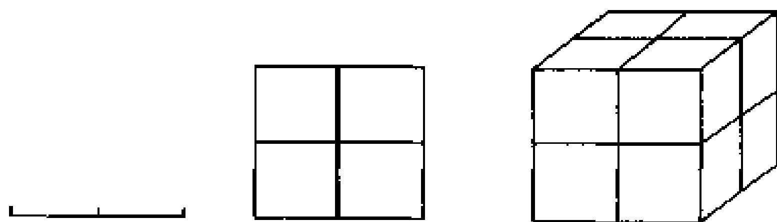


图8-3

将原边长分成3等分如何?

线段　　　分成 $3 = 3^1$ 个小线段;

正方形　　分成 $9 = 3^2$ 个小正方形;

立方体　　　分成 $27 = 3^3$ 个小立方体.

再一般些,边长缩小为 $1/a$,则

线段　　　分成 a^1 个小线段;

正方形　　　分成 a^2 个小正方形;

立方体　　　分成 a^3 个小立方体.

当某图形是由把全体缩小为 $1/a$ 的 a^D 个相似图形构成时,那么此指数就应具有维数的意义.一般称此维数为相似性维数.若根据这一维数,皮亚诺曲线的矛盾就可得以解决.从图 8-2 也可看出,皮亚诺曲线全体是由把它缩小成 1/2 的 4 个图形构成.因 $4 = 2^2$,皮亚诺曲线的相似性维数也为 2,与正方形的情况相一致.

相似性维数,虽然是再次构成的经验维数,但实际上它却具有经验维数所没有考虑到的性质.这从上述定义也可看出,相似性维数 D 可以不是整数.

定义　如果某图形是由把全体缩小成 $1/a$ 的 b 个相似形所组成,则相似性维数为

$$D = \frac{\log b}{\log a}(\Leftrightarrow b = a^D). \tag{1}$$

现在来回顾一下图 8-1 的科克曲线.如前所述,科克曲线是由把全体缩小成 1/3 时 4 个相似形构成的,所以,根据(1)式科克曲线的相似性维数可表示为

$$D = \frac{\log 4}{\log 3} = 1.2618\cdots, \tag{2}$$

这是一个非整数值.

取非整数值的维数,这对只熟悉经验维数的人来说,也可能会感到非常奇怪.但如果觉得科克曲线为 1 维有点太复杂,说它是 2 维(与皮亚诺曲线相比)又嫌太简单的话,那么 1.2618… 维也可能正合适.这个非整数值维数,恰好定量地表现了科克曲线的复杂程度.分形虽然一般都非常复杂,但其复杂程度却可用非整数维去定量化.如

果有 2 个不同的非整数维图形，一般维数高的图形更为复杂. 非整数维所起作用之大，通过下面的例子可窥知一斑.

虽然相似性维数是把经验维数扩大为非整数值的划时代的量，但原封不动使用这一定义，适用范围就非常局限了. 因为只有严密相似性的规则的分形，才能定义这个维数. 还存在能够使用于包括随机图形在内的任意图形的维数的定义. 例如豪斯道夫维数和盒维数，它们可以在任何集上定义，并且可以证明，它们与上面定义的维数相等. 这些定义超出了本讲座的范围，不再介绍.

8.1.6　几种基本的规则分形

1）康托尔粉尘与魔梯　康托尔粉尘，也称作康托尔集合，与科克曲线类似，是介绍分形时必定会出现的典型的分形. 这是 1883 年康托尔的创造物. 它可以说明分形的许多重要而有趣的性质，其应用非常广泛.

康托尔集这样构造：取长度为 1 的线段，不妨设它是$[0,1]$区间. 把它分成三等分，去掉中间的一份$\left(\frac{1}{3},\frac{2}{3}\right)$. 然后把剩下的两个线段$\left[0,\frac{1}{3}\right]$，$\left[\frac{2}{3},1\right]$再分成三等分，也去掉中间的一份$\left(\frac{1}{9},\frac{2}{9}\right)$，$\left(\frac{7}{9},\frac{8}{9}\right)$. 把这种做法一直继续下去，直到无限多的次数后所剩下的集合叫作康托尔集，或康托尔粉尘，如图 8－4 所示. 这是无限多个点分布在原来的线段上，并且点与点之间不会连接起来，点的分布也是疏密不均的.

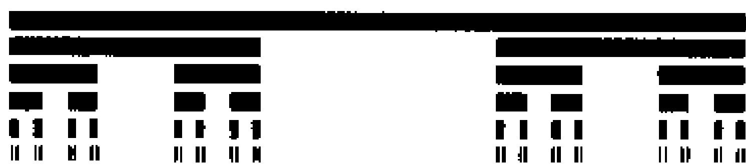


图 8－4　三分法康托尔尘粉的制造

第一次分为三份($a=3$),去掉一份,剩下的两份($b=2$)每份长度为 1/3. 按照前面的定义,康托尔粉尘的分数维是

$$D=\frac{\log 2}{\log 3}=0.6309\cdots.$$

不难算出,被去掉的区间的总长度是 1,等于开始时的区间长度. 所以在长度意义上讲,康托尔集的长度是 0. 康托尔集像尘埃,所以叫康托尔粉尘.

芒德布罗指出,康托尔集的结构过于规则,难以描述现实世界的自然事件. 不过可以想象,如果把点的分布模式相互间的关系随机化,就会得到与康托尔集有关的统计自相似的点集. 芒德布罗证明了,可以用它的维数描述电话传送线上的噪音分布的情况. 它还可以用以描述空气污染中尘云的结构.

现在用康托尔粉尘来造魔梯. 设想在[0,1]区间上均匀分布着某种物质,其密度为 1. 因为区间的长度为 1,所以总质量为 1. 假定在上述的分割过程中总的质量不变,而物质收缩,集中于区间 $\left[0,\frac{1}{3}\right]$和$\left[\frac{2}{3},1\right]$内,每段的长度为 1/3,质量为 0.5,所以密度在区间$\left[0,\frac{1}{3}\right]$和$\left[\frac{2}{3},1\right]$内为 3/2,在区间$\left(\frac{1}{3},\frac{2}{3}\right)$内为 0. 在第二次分割中物质进一步收缩,并假定它都集中在区间[0,1/9],[2/9,1/3],[2/3,7/9],[8/9,1],那么这些区间上的密度为$(3/2)^2$,而其它点上则为 0. 这个过程无限次地重复下去. 密度在康托尔集上无限增加,而在其它点上密度是 0. 如果设 $M(x)$ 表示$[0,x]$区间上的质量分布,则它的图象如图 8 - 5 所示. 若用 $\rho(x)$ 表示密度函数,则

$$M(x)=\int_0^x \rho(t)\mathrm{d}t.$$

注意到在空隙中没有物质,魔梯中便出现宽度各不相同的阶梯. 魔梯具有任意小的细微结构,且具有自相似性. 魔梯在物理学中很有用途,除代表质量分布外,还可以看作电荷、磁矩或某个现象的概率分布.

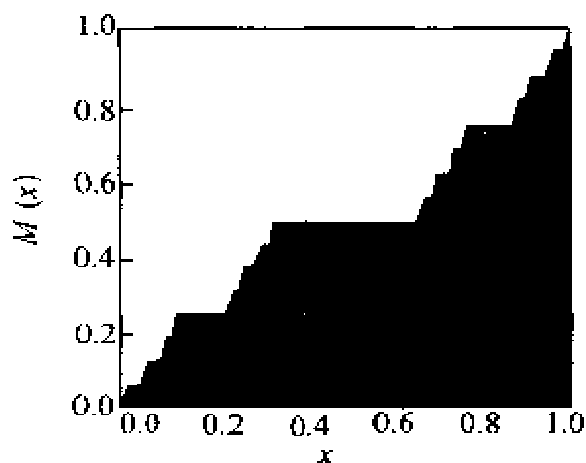


图 8-5

在结晶学上会遇到看上去象魔梯的晶体边界，这也与魔梯有一定联系.

2）谢尔平斯基(Sierpinski)衬垫和毯　1915 年，波兰数学家谢尔平斯基创造出了一种分形，后来这种分形就以他的名字命名为谢尔平斯基衬垫.衬垫的构造方法是这样的.从一个黑色等边三角形开始.第一步是把三角形的每一边长平分为二，并把分点连起来，于是原来的黑三角被分割成四个相同的小的黑三角，它们都是等边的，然后去掉中间的一个小的黑三角.按照这种方法，把留下的黑三角形继续这样处理，一直到无穷次数.所剩余的部分便微不可见.谢尔平斯基衬垫实质上是康托尔粉尘在二维空间中的一种广义的变体.图 8-6 表示产生谢尔平斯基衬垫的前四个阶段的图形.

在创造谢尔平斯基衬垫中，黑色三角形的边长每次都是平分为 2，而剩下的是三个相同的等边的小黑三角.由此，谢尔平斯基衬垫的维 $D=\ln 3/\ln 2=1.585$.在极限下，谢尔平斯基衬垫中的黑色部分完全看不清楚了，而成为一种典型的分形.

与谢尔平斯基衬垫相关的是谢尔平斯基毯片(图 8-7).其结构的开始为一正方形，然后把边长分成三等份，于是原正方块便分割成



图 8－6　谢尔斯基衬垫创作过程中的头四个阶段的图（从左到右）

9 个小正方块，去掉中央的一个，剩下 8 个. 以同样的办法无限地作下去，最后剩余下来的便是一条分形曲线了. 黑色部分在有限的阶段中还可以辨认，但在极限下，便变为零了，而白色方形边缘的总长度则是无限的. 依上例可以求得谢尔平斯基毯片的分形维 $D = \ln 8 / \ln 3 \approx 1.89$. 图 8 － 7 表示了前四阶段的谢尔平斯基毯片的构造.

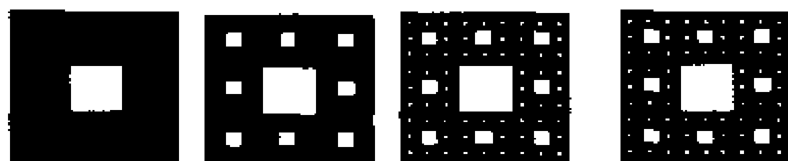


图 8 － 7　谢尔平斯基毯片头四阶段的结构

谢尔平斯基曲线在物理中是有用的. 例如在研究超导现象中，临界温度随磁场强度的变化便呈现为第十阶段的谢尔平斯基衬垫曲线. 在研究非晶态物质中，谢尔平斯基衬垫可以作为模型.

8.1.7　自然界中的分形

1）湍流　气体与液体统称为流体. 流体的流动形式分为两种：一种是层流，一种是湍流. 在层流中，流体的流动是平稳的，每个地方的流速与方向是均匀的，表现为类似层状的流动. 例如在圆管中呈层流状态的流体就好像分成了许多与管轴相平行的液层，与管壁接触的液层是不动的，而越接近中心的液层流速就越大. 这表明流体在层

流状态下基本保持沿管轴线的直线运动，而无横向运动.湍流与此不同.在湍流中，流体中每一点的流速与方向都不断变化，动量、热量、和质量的输送与交换比层流剧烈得多.由于流体微团间的激烈碰撞，加上聚合与破裂，湍流发声的强度也大大超过层流.风和河水的流动，看来似乎是平稳向前的，其实它们是湍流.河水不断出现大大小小的旋涡，二、三级的风也会出现小旋风.在自然界中，大部分流体的流动是湍流.从弥漫于房间中的香烟的烟呈非常复杂的现象也可以推断出那是湍流.

流动状态从层流向湍流的过渡过程叫转捩.流动转捩是雷诺(O. Reynolds)于1893年首先研究的.他发现圆管管流转捩主要受控于一个数学量，即雷诺常数 R.其计算公式是

$$R = \frac{\rho v L}{\mu},$$

其中 v 是流体的平均流速，L 是管道直径，ρ，μ 分别是流体密度和粘性系数.低于雷诺数的流体运动通常是层流，高于雷诺数的流体流动将出现湍流.

一般认为湍流具有分数维性质.空间湍流的分形维 D 大约为2.6.这只是一种结果.湍流中的许多问题包括它的分数维的确定仍在研究中.

地球的大气流动靠云才能看见.因为大气流动是湍流，所以云的形状和运动非常复杂.根据最近的观测确认云的形状为分形.

湍流的图样很容易制造出来.取一盆水，搅拌一下，然后滴一滴墨水进去，再覆盖一张纸在水面上，几秒钟后，把纸取出，于是纸上便显示出湍流图样，如图8-8所示.这个形状与云的形状相似.当然云的形状比它大几百万倍.

大气中的湍流是引起星光闪烁和遥远的光明灭不定的原因.大气的折射率取决于大气的密度与温度.它的折射率与大气的湍流有直接的关系，也就是说，星星的闪烁是由湍流形成的.

再看看我们的身体.在我们的身体中无时无刻不在保持各种生理流动,其中最重要的是血液循环和肺部呼吸.健康人体的血管和气管都具有良好的弹性,管壁可以吸收扰动能量,起着稳定的作用,因而生理流动的转捩雷诺数要远远超过工程中的刚性管流雷诺数.于是可以断言,正常人体循环系统的血液几乎保持着层状流动.在气管和支气管中气体的流动也是类似的.正常呼吸时,气体一直保持层流状态.唯当深呼吸和咳嗽时,才会发生湍流.上面谈的是健康人的生理流动.一旦循环系统和呼吸系统管道的弹性减弱,那么吸收扰动能量的能力就会大大减弱.这时就容易激发湍流,而且湍流旋涡会对病变的管壁造成进一步的损伤.湍流发声的强度要远大于层流,而且声音也有显著的差别.这就使得医生凭一对训练有素的耳朵和一只结构简单的听诊器听出许多病症来.湍流发声在医学听诊过程中起了这么重要的作用,真是出乎我们的意料.



图 8－8

2）河流　自然界中的河流也是一个典型的分形.因河流与它的分枝形状,不论从全体还是从支流来看都没有太大变化.借助主流长度与流域面积的经验关系,可算出河的主流的维数是1.2.日本名古屋大学的分数维研究会对河的分数维有详细的研究.根据他们的研究,世界各种河的主流的分数维维数在1.1～1.3左右.图8－9是亚玛逊河的形状.

3）肺和血管的构造　分数维比2大的曲面的表面积理论上可以任意大.能够很好地利用这一性质的组织是肺.肺从气管尖端成倍地反复分岔,使末端的表面积变得非常大.人肺的分数维大约为2.17.分数维越大使表面积变大的效率也越好,但这时曲面的凹凸

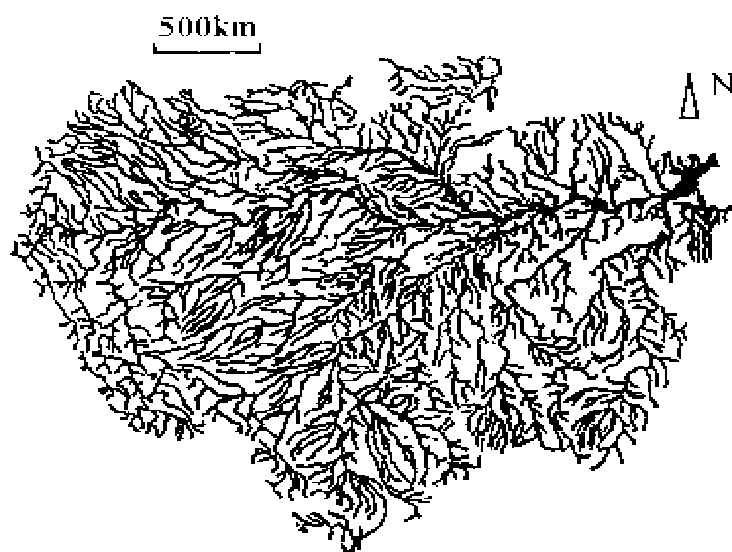


图 8-9 亚马孙河的形状

也变得更加厉害,这不利于空气的流通,为了兼顾起见才产生了2.17这一数值.

血管也呈分形结构,它必须把养分送到全身各个角落的细胞中.

视网膜血管的分形性质可能用于临床诊断.例如,正常人的视网膜血管与患糖尿病、高血压病的人的视网膜血管的分形维的值有差别,并且病情越重,差别越大.

除肺和血管外,动物体中还有不少组织是有分数维构造的.例如脑就是其中之一.人脑表面有各种不同大小的皱褶,它们是 2.73 2.79 维数构造.俗话说脑皱褶越多人越聪明,从分数维的角度看,可以这样说,脑的分数维维数越高就越有高维数的思考.

分形的意义在于探索自相似性,自相似性是跨越不同尺度的对称性.

§8.2 奇妙的混沌

8.2.1 混沌的定义

不论是在中国还是在西方，混沌的概念自古就有了．“易经”说：“混沌者，言万物相混成而未相离．”这可以看作古人给混沌下的定义．“西游记”一开头就说：

混沌未分天地乱，茫茫渺渺无人见．
白从盘古破鸿蒙，开辟从兹清浊辨．

指的是混沌先于宇宙，混沌孕育宇宙，宇宙出自混沌．

“辞海”上是这样定义混沌的：

亦作“浑沌”．1）古人想象中的世界开辟前的状态．《白虎通·天地》：“混沌相连，视之不见，听之不闻．”2）无知无识貌．

混沌一词的英文是 chaos．查看一下西方的辞典，例如朗曼英文辞典（LONGMAN DICTIONARY OF CONTEMPERARY ENGLISH）的注解在本质上与此完全一样，也是这两条．

到现代，混沌一词赋有了新的涵义，成为各个学科竞相关注的一个学术热点．英国皇家学会于 1986 年在伦敦召开的一次有影响的关于混沌的国际会议上，提出了下述定义：

3）数学上指在确定性系统中出现的随机状态．

在这个定义中又出现了两个玄妙的词：随机性与确定性．随机性指的是无定律，无规则，而确定性指的是受精确的，固定不变的定律的支配．这两个词联在一起不是一种悖论吗？诚然如此．通过后面的例子，我们就会知道，混沌完全是由定律支配的无定律状态．

研究混沌运动，探索复杂现象中的无序中的有序和有序中的无序，就是新兴混沌学的任务．我们将会看到，混沌无所不在．它存在于大气中，海洋湍流中，动物种群数的涨落中，股票价格的变动中，心脏的颤动中，……．世界是混沌的，混沌遍世界！

8.2.2 混沌的发现

法国大数学家庞加莱在 19 ~ 20 世纪之交研究天体力学，特别是三体问题时发现了混沌．但当时的数学家和物理学家都不理解，也不欣赏庞加莱的工作．原因是牛顿力学在科学中占有统治地位．从牛顿到庞加莱的二百年间的数学主要研究局部性，连续性，光滑性，有序性．这些经典理论用一层厚实而不易觉察的帷幕把混沌这块富饶的宝地给隔开了．庞加莱第一次在这道帷幕上撕了一条缝，暴露出后面有一大片未开垦的处女地．

8.2.3 蝴蝶效应

第二次世界大战期间，美国数学家洛仑兹成了一名空军气象预报员，这使他迷上了天气预报．60 年代，他开始用计算机模拟天气情况．在对气象预报的研究中，他发现了天气变化的非周期性和不可预报之间的联系．在天气模型中他看到了比随机性更多的东西，看到了一种细致的几何结构，发现了天气演变对初值的敏感依赖性．洛仑兹给了一个形象的比喻："巴西的一只蝴蝶扇动几下翅膀，可能会改变 3 个月后美国得克萨斯的气候"．这被称为"蝴蝶效应"．用混沌学的术语来表述就是，系统的长期行为对初值的敏感依赖性．

中国有句成语"差之毫厘，失之千里"，说的就是这个意思．这种对初值的敏感性的例子在日常生活中并不少见．例如，一个考生晚了两分钟离开家门，误了一趟班车，因迟到而考砸了一门课，致使高考落榜．对"紧要处"的敏感依赖，对个人而言，可导致截然不同的人生结局；对于国家，可导致兴盛或灭亡．英国一首民歌是这样写的：

丢失一个钉子，坏了一只蹄铁；坏了一只蹄铁，折了一匹战马；
折了一匹战马，伤了一个骑士；伤了一个骑士，输了一场战斗；
输了一场战斗，亡了一个国家．

这是说事件经过逐级放大会导致严重后果．为了观察这种现象，你不妨自己做个实验．把一片树叶放入潺潺的小溪中，然后再把另一

片树叶精确地放入与前一片树叶相同的地方.刚开始,两片树叶的运动可能会一样,但不久它们所表现出来的运动形式会截然不同.原因之一就是你把第二片树叶放入小溪的地方不可能与第一片树叶完全相同.这点微小的差异会逐渐放大,最终表现出完全不同的行为.

8.2.4 线性与非线性

线性过程指的是这样的过程,如果初始时刻任一变量的一点变化会使得以后的时间内这个变量或其它变量也产生一点变化;初始时刻2倍大的变化会使得以后的同样时间内也产生2倍大的变化.2倍也可换为5倍或0.5倍.我们最熟悉的线性过程是购物.例如买柿子,一块钱买一个,十块钱就买十个.

在数学上,线性过程用一次方程来描述.如

$v = kt$,k 是常数.

$y = ax + b$,a,b 是常数.

$w = ax + by + cz$,a,b,c 是常数.

由于一次方程在平面上的图形是直线,所以用一次方程表示的过程就叫作线性过程.下面的方程是非线性方程的例子:

$$w = z^2 + c,\ c\ \text{是常数}; \tag{1}$$

$$F = kx + \varepsilon x^3,\ k,\varepsilon\ \text{是常数}. \tag{2}$$

后一方程是非线性弹簧的恢复力 F 的表示式.研究非线性问题的理论模型称作非线性系统,通常用非线性方程来描述.

线性过程是容易处理的过程.例如,在弹性限度以内,弹簧的恢复力 F 就与位移 x 成正比(在(2)式中,$\epsilon = 0$):

$$F = kx.$$

线性过程所以重要有两个原因:第一,许多实际的现象在所限制的时间内和限制的变量范围内近似可看成是线性的,所以通常的线性代数模式能够模拟它们的行为.例如,一个摆动角度很小的单摆可近似看作是线性系统.第二,线性方程可以用许多方法处理,而这些方法对于非线性方程却是无能为力的.

丰富多彩的大千世界是非线性的.一般来说,一个和尚挑一担水,三个和尚挑三担水的理想情况是不会出现的.常常出现的情况是:"一个和尚挑水吃,两个和尚抬水吃,三个和尚没水吃."长期以来人们把注意力集中于线性问题,是出于无奈,因为非线性问题太困难了,缺乏处理它的手段.近几十年来情况发生了很大的变化,特别是计算机的诞生,使研究非线性有了更得力的工具.

线性过程太简单了,太单调了,它不会产生混沌.任何混沌系统必然是非线性的.但这里还必须指出,虽然混沌要求非线性,但非线性并不保证有混沌.

下面举一个产生混沌的例子.

8.2.5 函数的迭代

1) 函数的迭代 我们需要对函数引进一种新运算,称为函数的迭代运算.这种运算过去不大考虑.先看一个例子.设 $s(x)=\sqrt{x}$,那么

$$s(256)=\sqrt{256}=16,$$
$$s(16)=\sqrt{16}=4,$$
$$s(4)=\sqrt{4}=2,$$
$$\cdots\cdots$$

这是反复作开方运算.这就是一种函数的迭代运算.现在看一般情形.

给定了一个函数 $f(t)$,考虑函数 f 到自身的多次复合运算,并记为:

$$f^1(t)=f(t)$$
$$f^2(t)=f(f(t))=f\circ f(t)$$
$$\cdots$$
$$f^n(t)=f\circ f^{n-1}(t).$$

注意,f^n 在这里不是 f 的 n 次方幂,f 的 n 次方幂用记号$[f(t)]^n$来表

示.用 f^0 表示恒同映射,即

$$f^0(t) \equiv t.$$

例如,若 $f(t) \equiv 2t + 1$,则

$$f^0(t) \equiv t, f^1(t) \equiv 2t + 1,$$

$$f^2(t) = f \circ f(t) = f(f(t)) = 2f(t) + 1 = 2(2t + 1) + 1$$
$$= 4t + 2 + 1 = 4t + 3.$$

线性函数的迭代不会产生出有趣的结果.在有理函数中除去线性函数外,最简单的函数是二次函数:

$$f(x) = ax^2 + by + c.$$

这个函数的迭代就会产生出许多有趣的结果.

函数迭代生成的系统称为动力系统.如果变量 t 取复值,所产生的系统称为复动力系统,这是当前复分析研究的重要方向之一.

2) 不动点与周期点

定义　若 $f(t_0) = t_0$,则称 t_0 是 f 的一个不动点.

例　设 $f(t) = \sin t$.由 $\sin 0 = 0$ 知,0 是 $\sin t$ 的一个不动点.

设 $f(t) = \frac{1}{2}t(t-1)$.由 $f(3) = 3$ 知,3 是 $f(t)$ 的一个不动点.

定义　若 $f^p(t_0) = t_0$,则称 t_0 是 f 的一个周期点,p 称为周期点的阶数.

例　设 $f(t) = \frac{1}{t}$.取 $t_0 \neq 0$,易见,t_0 是一个 2 阶周期点.

动力系统中最基本的概念之一是轨道及其收敛性.对任一实数 t_0,定义 $Q^+(t_0)$ 为正向轨道,它是集

$$Q^+(t_0) = \{t_0 = f^0(t_0), t_1 = f^1(t_0),$$
$$t_2 = f^2(t_0), \cdots, t_n = f^n(t_0), \cdots\}.$$

轨道一般是一个无限集.

如果在 $Q^+(t_0)$ 中存在一个最小的 $p > 0$,使得 $f^p(t_0) = t_0$,则 t_0 是一个周期点.这时 $Q^+(t_0)$ 就由周期循环组成,记为

$$Q^+(t_0)=\{t_0=f^0(t_0),t_1=f^1(t_0),\\ t_2=f^2(t_0),\cdots,t_{p-1}=f^{p-1}(t_0)\}.$$

换句话说,这 p 个点满足这样的关系

$$t_1=f(t_0),t_2=f(t_1),t_3=f(t_2),\cdots,t_p=f(t_{p-1}).$$

特别地,当 t_0 为不动点时.对应的正向轨道为

$$Q^+(t_0)=\{t_0\}.$$

当 $Q^+(t_0)$ 是无限集的时候,出现三种情况:$Q^+(t_0)$ 有有限极限;$Q^+(t_0)$ 趋于 ∞;$Q^+(t_0)$ 无规律.

8.2.6 人口模型

迭代的起源之一是人口问题,所以我们从人口模型谈起.假定在开始时刻某地的人口总数是 x_0,n 年后的人口总数是 x_n,那么第 $n+1$ 年的人口增长率是

$$r=\frac{x_{n+1}-x_n}{x_n}.$$

如果该地人口的增长率是常数,即 r 为常数,则上面的方程将对所有的 n 都成立.把它改写为线性方程

$$x_{n+1}=(1+r)x_n=f(x_n);f(x)=(1+r)x.$$

于是

$$x_n=(1+r)x_{n-1}=(1+r)^2x_{n-2}=\cdots=(1+r)^nx_0,$$

即

$$x_n=(1+r)^nx_0.$$

这就是说,人口的增长是指数增长.不仅人口的增长模型如此,动物的繁殖,细菌的繁殖在一定的时间间隔中都是指数增长.这是一个典型的线性动力系统的模型.

但是当时间间隔太长时,这个模型将不符合实际情况.因为食品有限,人类的生存空间也有限,人口的无限增长是不可能的.荷兰的数学生物学家弗尔哈斯特(Verhulst)在 1837 年指出,在一段时间以

后人口的增长将达到极限 X. 当人口总量接近极限 X 时，人口增长率将从 r 降到 0. 在数学上处理的办法是，令 $p_n = \frac{x_n}{X}$，从而 $0 \leqslant p_n \leqslant 1$. 动态方程代之以

$$p_{n+1} = kp_n(1 - p_n).$$

它是二次函数

$$y = f(x) = kx(1 - x).$$

迭代的结果. 由 $f(x)$ 所实现的映射称为逻辑斯蒂映射.

从数学上看，逻辑斯蒂映射是简单的，但它却是产生混沌的一个范例. 不久我们就会看到，这个简单的方程孕育着数学中可能有的最复杂、最优美的性态. 它告诉我们简单的系统可能产生复杂的行为. 也就是说，激烈的变化不一定有激烈的原因. 研究这个简单系统是帮助我们理解复杂的系统，如天气预报、股票涨落等的钥匙.

8.2.7　逻辑斯蒂映射

现在来研究逻辑斯蒂映射.

1) 几何意义　函数

$$y = kx(1 - x)$$

的图象是开口向下的抛物线，如图 8 − 10 所示. 由

$$y = kx(1 - x) = -k\left(x - \frac{1}{2}\right)^2 + \frac{k}{4}$$

知道，当 $x = 1/2$ 时，y 取到最大值 $k/4$. 抛物线关于直线 $x = 1/2$ 对称.

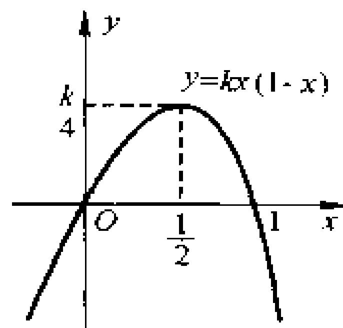


图 8 − 10

下面我们总假定 k 是 0 到 4 间的常数，x 在区间[0,1]上取值.

对给定的 k，逻辑斯蒂映射是一个以不均匀的方式拉长或压缩[0,1]区间. 例如取 $k = 4$，这时

$$y = 4x(1 - x).$$

当 x 从 0 变到 1/2 时，y 从 0 变到 1. 这个映射把[0,1/2]拉长成[0,1]. 当 x 从 1/2 变到 1 时，y 从 1 变到 0. 这又是拉长，但将原区间颠倒了过来. 映射的总效果是拉长线段[0,1]使它覆盖[0,1]线段两次.

如果把 k 取成 3，效果还是拉长与折叠. 不过是把[0,1/2]拉长成[0,0.75]而已. 当 $k<2$ 时，映射的效果是压缩与折叠. 如果 $k>4$，则迭代的结果将超出[0,1]区间. 多次迭代的结果，会使有的 x 值很快趋于无穷. 我们先假定 k 在区间[0,4]上变化.

要研究逻辑斯蒂映射的动力学性质，必须考虑它的长期形态，即多次反复迭代的结果. 参数 k 起着重要的作用. 不同的 k 值会使逻辑斯蒂映射呈现出不同的性态：定态，周期循环和混沌.

2）定态区间 当 $k\in[0,3]$ 时，从动力学观点看，逻辑斯蒂映射是最不重要的，这是定态区域. 先看两个实例.

当 $k=2$ 时，$f(x)=2x(1-x)$. 取 $x_0=0.1$，反复迭代得如下结果：

$$x_0=0.1,\quad x_1=0.18,\quad x_2=0.2952,$$
$$x_3=0.4161,\quad x_4=0.4859,\quad x_5=0.4996,$$
$$x_6=0.4999,\quad x_7=0.5,\quad x_8=0.5.$$

从 x_7 之后，迭代就停止在 0.5. 所以，我们就说在 $x=0.5$ 处有一个吸引子. 这是一个稳定的状态，称它为定态.

取其它的初始值，例如 $x_0=0.2,0.3,0.6$，等，作迭代运算，所得结果也一样，总以 0.5 告终. 事实上，再取 $x_0=0.2$，得

$$x_1=0.32,\quad x_2=0.4352,\quad x_3=0.4916,$$
$$x_4=0.4940,\quad x_5=0.4999,\quad x_6=0.5,\cdots.$$

我们指出，$x=0.5$ 是映射的不动点：

$$f(0.5)=2\times0.5(1-0.5)=0.5.$$

这个结果也可从数学经济学家经常使用的一种叫蛛网图的图象上看

到,如图8－11所示.在图上画出直线 $y=x$,它与 $y=kx(1-x)$ 交于点 $\left(1-\frac{1}{k},1-\frac{1}{k}\right)$.设 x_0 是初始值.从 x_0 开始,画出垂直线 $x=x_0$,这条垂直线与抛物线有一个交点,这点的纵坐标是 $x_1=kx_0(1-x_0)$.再从这个交点画水平线 $y=x_1$,这条水平线与直线 $y=x$ 交于一点,其坐标为 (x_1,x_1).交点的横坐标为 x_1.从点 (x_1,x_1) 出发作垂直线 $x=x_1$,与抛物线交出第二个交点,这点的纵坐标是 $x_2=kx_1(1-x_1)$.依此反复作下去.在抛物线与直线 $y=x$ 之间形成一个个"阶梯".这些"阶梯"沿斜线上行,然后向内盘旋而达到不动点.

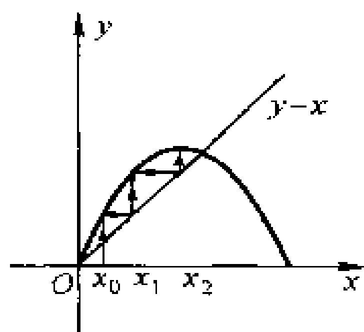


图8－11

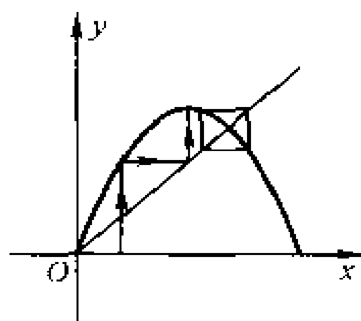


图8－12

只要 $k<3$,蛛网总是向内盘旋而达到不动点.对0到3范围内的 k,得到的总是单个稳定的不动点.但是当 $k>3$ 时,情况就不同了.

3）周期性态　当 $k>3$ 时,逻辑斯蒂映射迭代的结果是什么呢?取 $k=3.2$ 试试看.这时,$f(x)=3.2x(1-x)$.取 $x_0=0.5$,反复迭代,得如下结果(取4位小数):

$$x_1=0.8,\quad x_2=0.512,\quad x_3=0.7995,$$
$$x_4=0.5130,\quad x_5=0.7995,\quad x_6=0.5130,$$
$$x_7=0.7995,\quad x_8=0.5130,\cdots.$$

从 x_3 以后出现循环，x 交替地取 0.7995 和 0.5130 两个值．迭代的蛛网图如图 8－12 所示．它收敛到一个正方形的环上，x_i 的值交替地取 0.7995 和 0.5310 两个值．这是一个周期为 2 的循环．

取其它的初始值，例如 $x_0 = 0.2, 0.3, 0.6$，等，作迭代运算，所得结果也一样，也得到周期为 2 的循环．如果你有一台计算机，可以编一个程序自己算算看．取别的初始值作迭代最终也会得到这一结果，但收敛速度可能不一样．

k 再增大，结果如何呢？取 $k = 3.5$ 算一算．这时，

$$y = 3.5x(1 - x)$$

令 $x_0 = 0.5$ 作迭代，得如下结果：

0.5，0.875，0.3828，0.8270，0.5007，0.875，0.3828，0.8270，…．出现了周期为 4 的循环．取其它的初始值，例如 $x_0 = 0.2, 0.3, 0.6$，等，作迭代运算，所得结果也一样，仍出现周期为 4 的循环．

取 $k = 3.56$，对 $y = 3.56x(1 - x)$ 作迭代，将出现周期为 8 的循环．k 再增大会出现什么情况呢？还是周期性态吗？

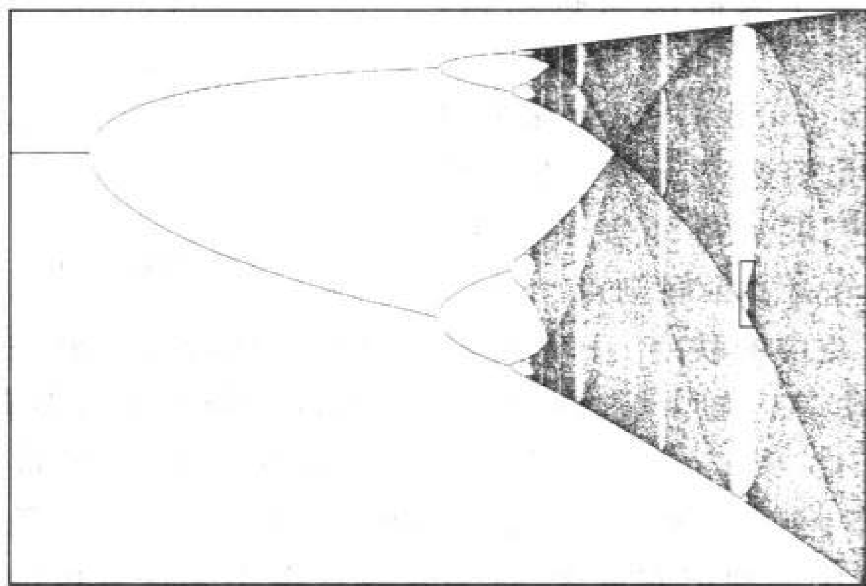


图 8－13

4) 混沌区域　当把 k 的值增加到 3.5 左右时，周期 2 吸引子失稳，出现周期 4 的循环. 当把 k 的值增加到 3.56 时，周期又翻到 8. 当 k 的值增加到 3.567 时，周期达到 16. 此后周期迅速加倍：32，64，128，⋯.

当 k 的值增加到 3.58 左右时，逻辑斯蒂映射变成混沌.

我们还可以画另一张图(图 8 - 13)来展示上述的迭代过程与参数 k 的关系. 我们把这张图叫作逻辑斯蒂映射的分岔图. 借助分岔图可以使我们对逻辑斯蒂映射的动力学性质有一个整体认识. 取 k 为横坐标，x 为纵坐标. 对每个 k 找出 k 所对应的吸引子，也就是相应的周期点，这样我们就得出图 8 - 13 所示的分岔图. 在 0 到 1 的区间内，每一条竖直线与图象的交点就是吸引子. 例如当 $k < 3$ 时，仅有一个交点，因此也只有一个吸引子.

$k = 3$ 时，一条曲线分成两条，因而这里出现一个分岔. 随着 k 的增加，分岔加倍，再加倍，⋯. 你会看到一个美丽的树状结构，我们把它叫作无花果树(图 8 - 14). 在 $k = 3.58$ 附近，无花果树终于分成无穷多个分支. 无花果树的分支扩展为混沌吸引子带. 分岔图上布满了无规则点.

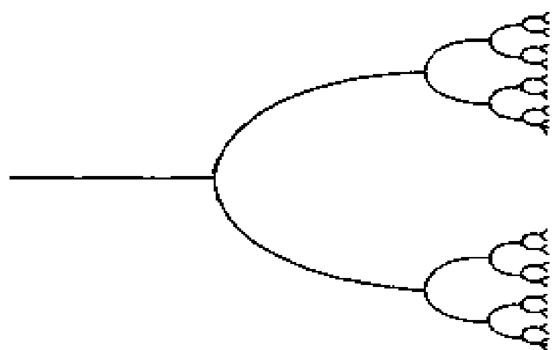


图 8 - 14

5) 混沌中的秩序　对 $k = 3.58$ 以上的混沌区域进行仔细地分析会揭示出惊人的事情：在混沌行为的背后隐藏着许多有趣的现象. 在图 8 - 13 中，有一些细长白条，这些白条构成周期窗口. 在 $k = 3.835$ 附近包含三个小无花果树(图 8 - 15). 为了显示出细节部分，选取其中一棵把它放大. 你会发现这棵小无花果树也以混沌带告终. 这个混沌带中又有细长白条. 这就是说，窗口中有窗口. 小窗口中又含有更小的无花果树. 这个过程可以越变越小地继续下去.

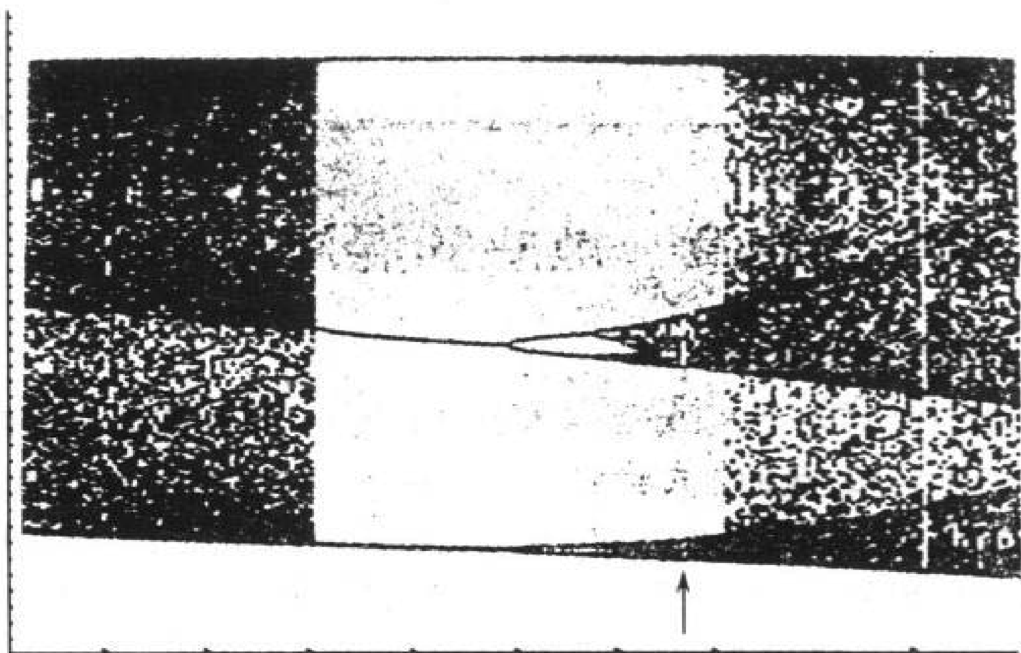


图 8 - 15

实际上,每一个窗口都是整个图形的小的复制品,包含完全相同的自身缩影.这就是自相似性,典型的分形行为!

这是一个很有启发性的例子.它至少给我们两点启示:1)确定的方程可以有混乱的输出;2)混沌与分形奇妙地联系了起来.

8.2.8 茹利亚集

从本世纪20年代起,法国数学家茹利亚(G.Julia)和法都(M.P.Fatou)开始研究复动力系统.例如,他们研究了形如

$$w = z^2 + c$$

的映射,其中 z 和 c 都是复数.这个映射可以看作逻辑斯蒂映射的复模拟.c 是复参数,相当于逻辑斯蒂映射的参数 k.现在的问题是,固定 c 值不变,从一个给定的初始值 z 出发,利用上面的公式作迭代,看发生什么情况.

首先考察最简单的情况:$c = 0$.设 z_0 为初始值,这时动力规律为

$$z_n = z_{n-1}^2 = \cdots = z_0^{2^n}.$$

当 $|z_0| < 1$ 时，$z_n \to 0$. 这就是说，原点 0 是系统的吸引子. 当 $|z_0| > 1$ 时，$z_n \to \infty$. 这表明，∞ 也是系统的吸引子. 在 $c = 0$ 的情况下系统只有两个吸引子. 整个复平面分成两个区域，一个区域以 0 为吸引子，另一个区域以 ∞ 为吸引子. 当 $|z| = 1$ 时，

$$|z_n| = |z_0^{2^n}| = |z_0|^{2^n} = 1.$$

这时迭代序列的点始终在单位圆周 $|z| = 1$ 上.

在这个例子中，复平面被一条边界曲线 — 单位圆周分为两个不同的吸引子区域，我们把这条曲线叫作尤利亚集. 当 c 的值变化时，尤利亚集呈现异常复杂的状况，且无比美丽. 尤利亚和法都都知道它们难以置信地复杂，但他们都无法看到真实图形. 借助现代计算机，现在不难把它们画出来. 其中有海马形、兔形、甲虫形、宇宙尘形、玩具风车形，等等，花样层出不穷.

当 $c = 0.31 + 0.04\mathrm{i}$ 时，有限吸引子由一点组成. 它和无限吸引子的边界是图 8 - 16 所示的漂亮的分形圆. 如果你任取边界的一部分进行仔细观察，你就会发现熟悉的、无限重复的自相似现象. 这正是分形曲线的特征.

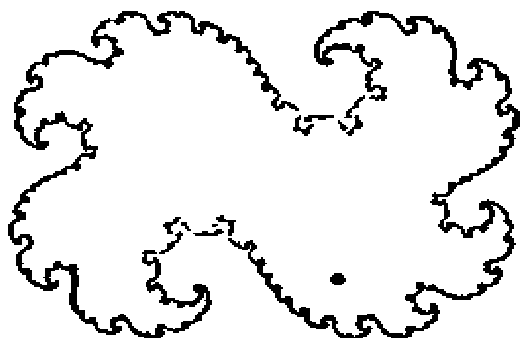


图 8 - 16

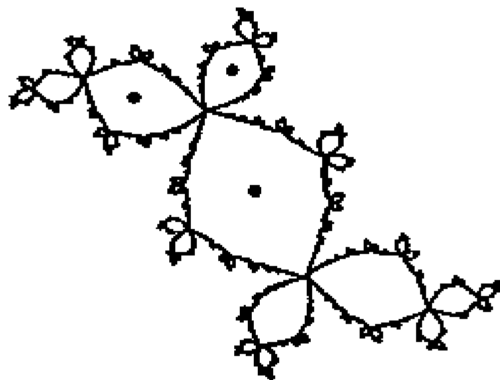


图 8 - 17

虽然只有计算机的出现才使人们有可能深入研究这样的图形，但尤利亚和法都两人早已证明了：这类图形的任何一段边界，不管它多么小，都包含了整个曲线所需的全部信息．正是为了纪念茹利亚才把这样的边界集叫作茹利亚集．

当 $c=-0.12+0.74i$ 时，茹利亚集由图 8－17 表示．有限吸引子由三点循环构成．图 8－18 给出了另一些尤利亚集的例子，其中有的区域退化为“尘点”或“树枝”．通过参数 c 的不同选择，茹利亚集展示出丰富多彩的结构．有的茹利亚集是连成一体，有的是一盘散沙．这说明了参数 c 的重要性．

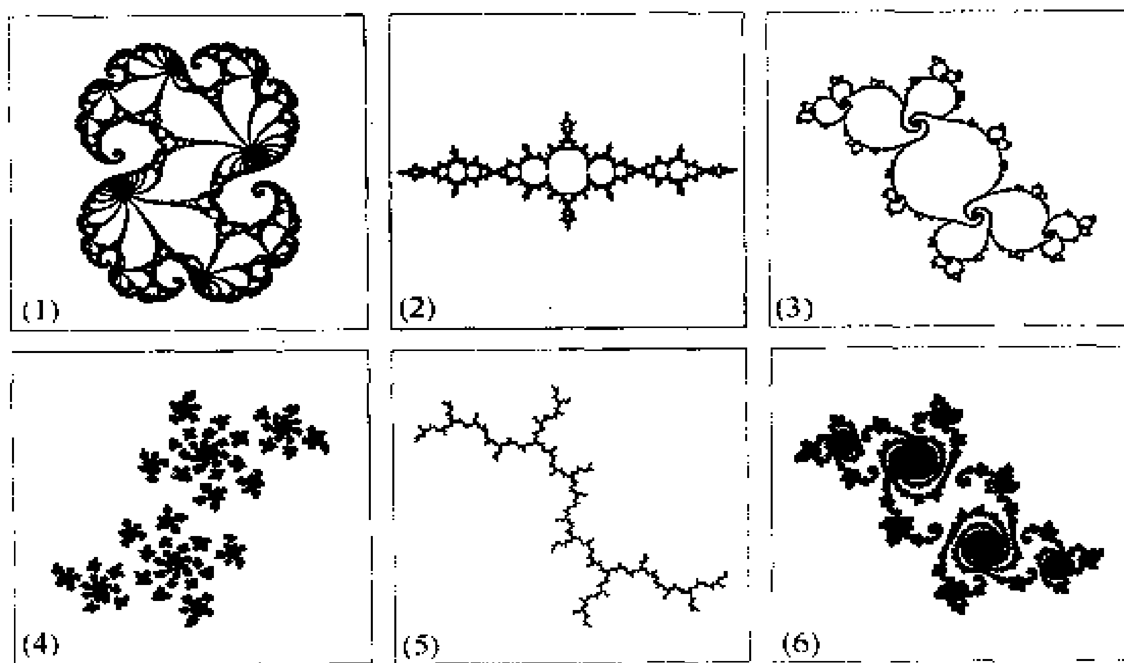


图 8－18

8.2.9 芒德布罗集

现在我们把复映射 $f(z)=z^2+c$ 与逻辑斯蒂映射 $f(x)=kx(1-x)$ 作一比较．我们看到，x 与 z 的作用相似，k 与 c 的作用相

似．每一个 c 有它自己的茹利亚集；这与每一个 k 有它自己的吸引子的事实相类似．对逻辑斯蒂映射我们找到了一种图形，它不但能表示对给定的 k 来说吸引子是什么，而且能显示出吸引子如何随 k 的值变化，这就是分岔图．它还使我们得到一个美妙的无花果树．对给定的 c 值也有一个类似的图形，它可以展示茹利亚集如何随 c 在复平面上变化而变化．这次我们得到的不是无花果树，而是甲虫状集，这就是芒德布罗集(图 8 － 19)．

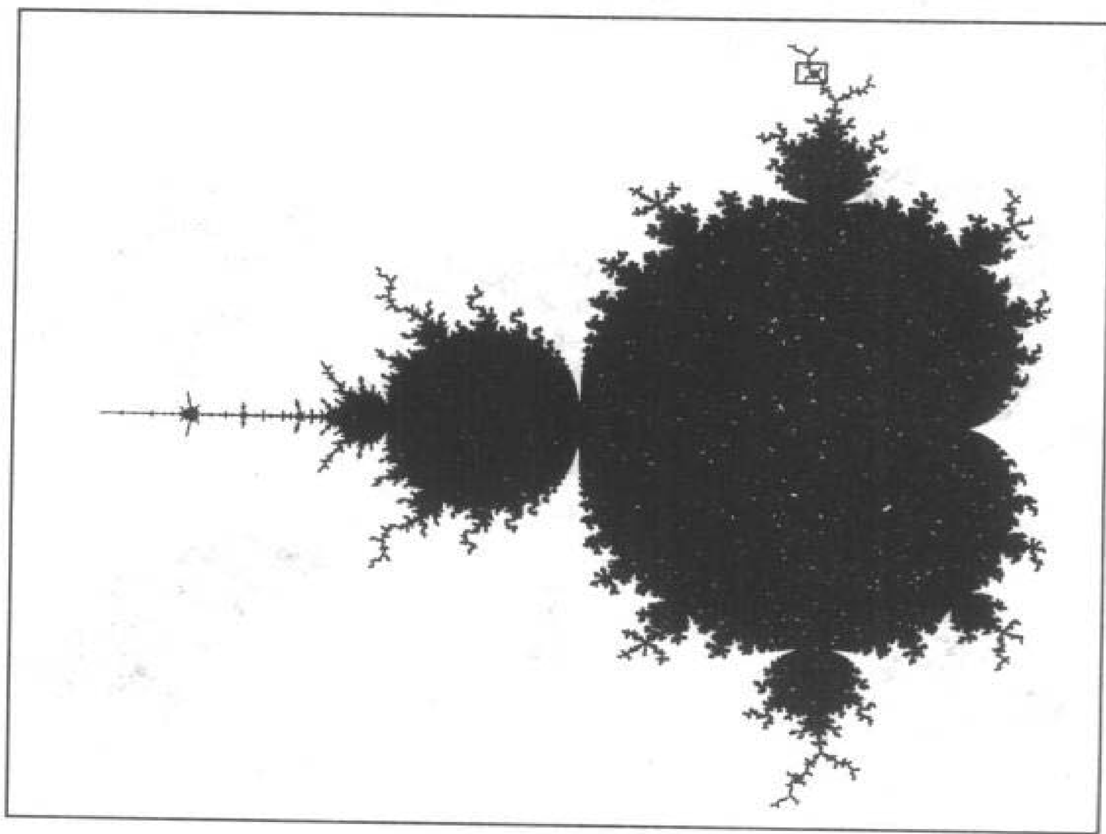


图 8 － 19

现在我们来制作芒德布罗集．在复平面上任取一点 c，对一切可能的 z 用 $f(z) = z^2 + c$ 作迭代映射，并对 c 求茹利亚集．看看它是否是连通的，如果是连通的就把 c 涂黑；如果不连通就把 c 途成白色．对复平面上的每一个 c 都这样做．结果我们就得到甲虫状集，如图 8 － 19 所示．

芒德布罗集最突出的特征之一是它的自相似性.在芒德布罗集很深的地方,大约百万分之一大小的地方,你会发现小甲虫状集(图8－20).每一个细节都完整无缺,它也有自己的小甲虫状集.就像逻辑斯

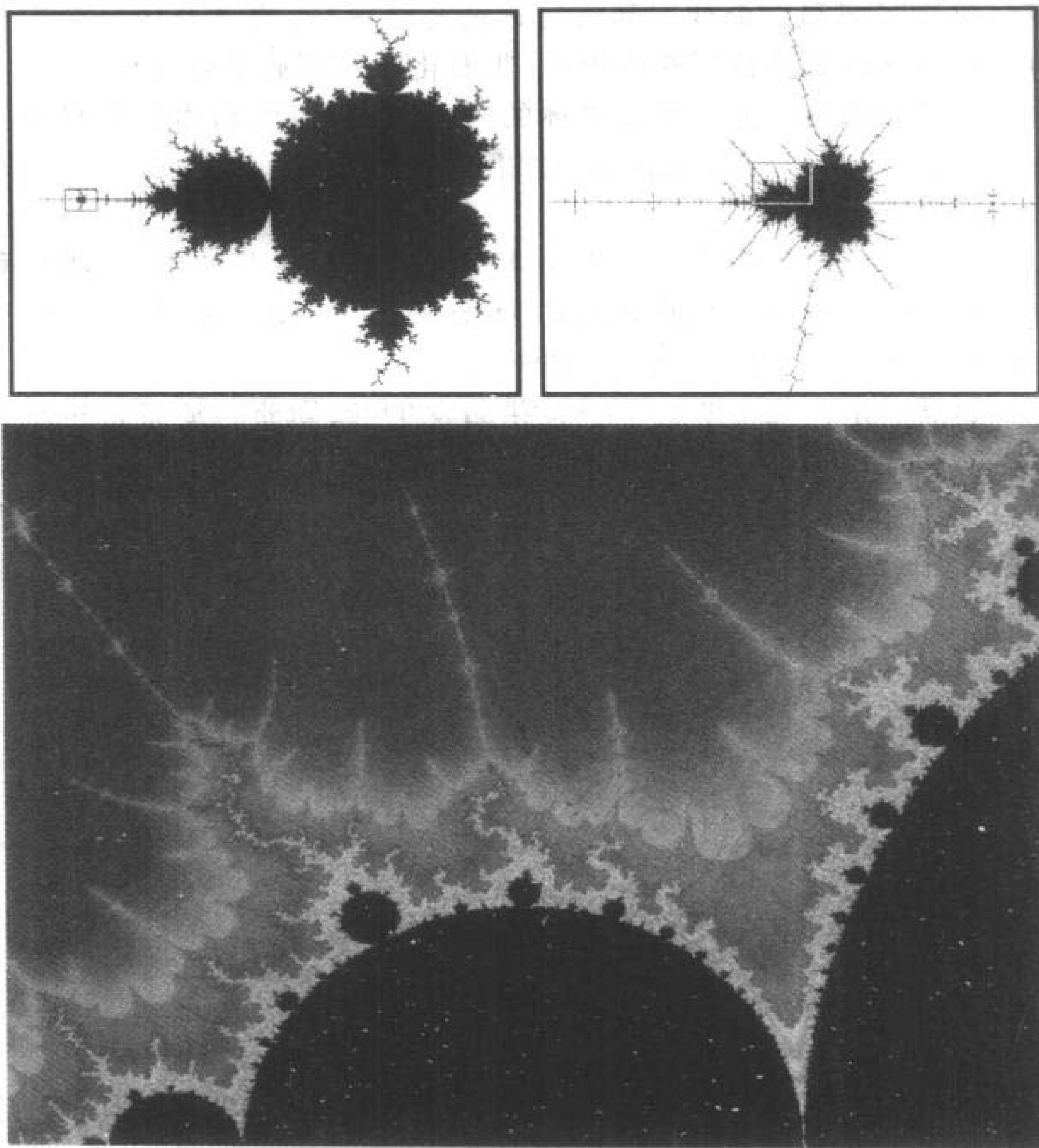


图8－20

蒂映射的分岔集具有自己的完整复本的窗口一样，芒德布罗集亦然.

芒德布罗集与茹利亚集的具体关系是什么呢?

c 的不同位置决定了不同的茹利亚集.或者将平面分成一个或多个内部区域和一个伸展到无穷远的外部区域，或者产生一个退化为没有内部区域的茹利亚边界集.非退化的尤利亚集有四种.

如果 c 选自芒德布罗集主体的内部，则相应的动力系统只有一个吸引子，即映射的不动点.这时茹利亚集是一个分形变形圆，如图 8－16 所示.

如果 c 选自与芒德布罗集主体相连的某个苞芽的内部，则茹利亚集由无限多个分形变形圆组成.如图 8－17 所展示的是，c 取自芒德布罗集顶部(左边的)大苞芽的中心.

如果 c 是芒德布罗集一个苞芽的发生点，则茹利亚集将呈现许多卷须，如图 8－18 的(1)或(2)所示.

最后，如果 c 是芒德布罗集的边界点，则茹利亚集就是著名的济格盘.图 8－21 是济格盘(siegel)的一个例子.这里存在一个不变

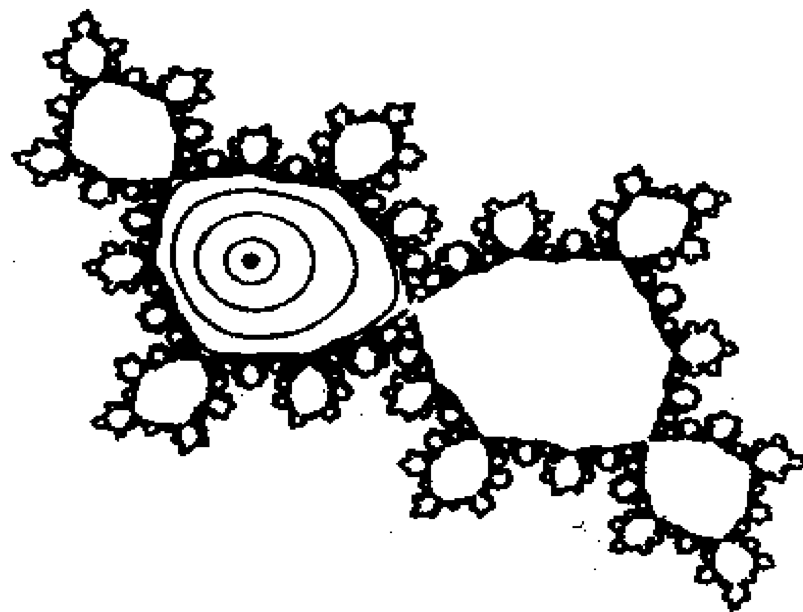


图 8－21

圆环绕的不动点．在这种情况下，茹利亚集所包围的区域的内点将逐渐地趋向包含不动点的圆盘．

非退化的茹利亚集就是上面四种．那么，其它情况呢？芒德布罗集的放大图表明它被一些发状分岔触角所包围．如果 c 正好选在这些触角之一，那么就得到形状类似的茹利亚集．图 8 － 18(5) 所表示的是 $c = i$ 时的例子．这时系统只有一个吸引子，即无穷吸引子．除去茹利亚集本身的点外，其它点将都变到无穷远去．

现在只剩下一种可能了，即 c 选自芒德布罗集的外部．这时无穷远是唯一的吸引子，而茹利亚集被分解成一些称为法都(Fatou dust)尘的一堆点，如图 8 － 18(4) 所示，c 离芒德布罗集越远，这些尘点就变得越细．这些尘点的图案呈现分形结构．

芒德布罗集的边界在动力系统中扮演着重要的角色，引起人们的巨大兴趣．边界本身具有复杂的分形性质．在蒙德尔布罗集的边缘上选取一点把它放大．在新图形的边缘上再取一点，再放大，得到图 8 － 20 所示的图形．这种放大过程还可以不断继续下去．每一次放大都展现出新颖而使人惊奇的结构．旋涡、海马、仙人掌、细蛇、虫状斑点，等等．

芒德布罗集与所有的动力系统都有密切的联系，它在数学中占有一个特殊而基本的地位，就像圆和多边形在经典数学中所占有的地位一样．

上面讨论的动力系统反映了自然界的一种普遍模式，对近代科学的发展产生了巨大影响．这种模式应当尽早在中等教育中出现．著名的生物数学家罗伯特·梅(Robert M May) 在 1976 年就指出："这种学习不像微积分那样涉及到那么多的复杂概念，这会使学生大大丰富对非线性系统的直觉认识．"

第九章　一笔画和邮递路线问题

社会十分尊重数学，这可能不是因为这个学科的内在美，而是因为数学是社会极其需要的一种艺术.

L. Bers

9.1.1　问题的提出

这一讲讨论一个来自拓扑学的问题，这就是一笔画和邮递路线问题，一个很有实用价值的问题.问题是这样提出的：

一个邮递员送信，每次都要走遍他所负责的投递范围的每一条街道，完成任务后回到邮局.问，他沿怎样的路线走，所走的路程最短？

这个问题是邮递员每天都要碰到的问题，叫作最短邮递路线问题，是我们中国人最先提出的.这个问题具有普遍意义.还有许多其它问题，在实质上也属于邮递路线问题.例如铁路检察员，他必须走过每一条铁路一次，然后回到出发地.最理想的邮递路线当然是从邮局出发，走遍邮递员所辖区域的每一条街，而且每条街只走过一次，最后回到邮局.这样的路线显然是最短的，因为没有重复.问题是，这样的路线能找到吗？先看几个例子.

例　为图 9 - 1 所示的街道图找出最短的邮递路线.假定邮局位于点 A.

解　可以找到最短的邮递路线，如图 9 - 2 所示.

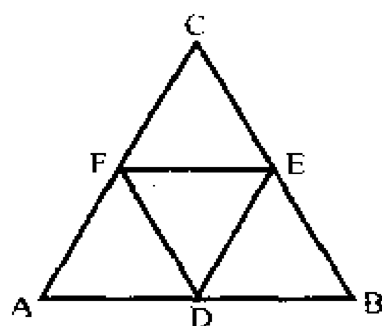


图 9-1

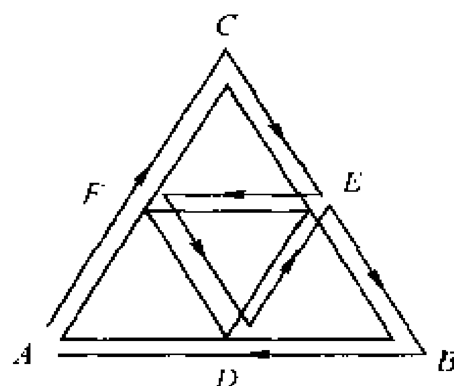


图 9-2

例 为图 9-3 所示的街道图找出最短的邮递路线.假定邮局位于点 A.

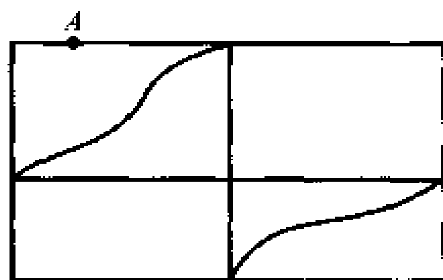


图 9-3

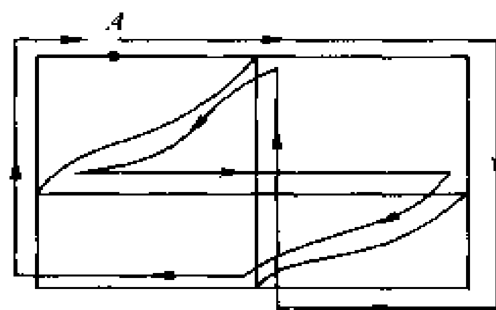


图 9-4

解 可以找到最短的邮递路线,如图 9-4 所示.另一条最短的邮递路线如图 9-5 所示.还有一条最短的邮递路线如图 9-6 所示.可见在有的问题中解法不惟一.

例 为图 9-7 所示的街道图找出最短的邮递路线.假定邮局位于点 A.

解 最短的邮递路线不存在.不管你怎么走,不是有的街道没有走到,就是有的街道要重复走.原因在什么地方?BC 是死胡同,进

去后还得沿原路出来.

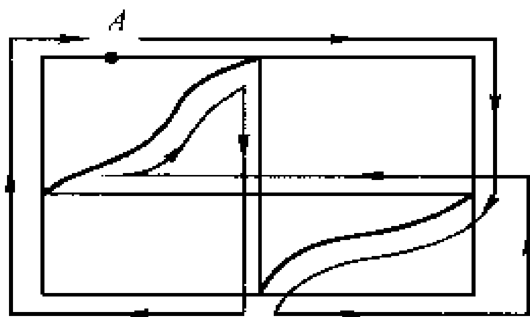


图 9 - 5

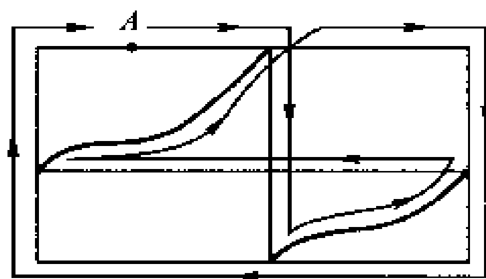


图 9 - 6

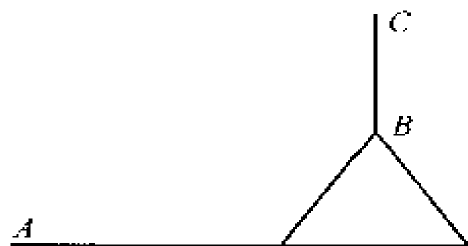


图 9 - 7

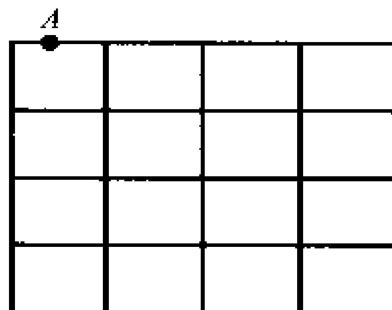


图 9 - 8

例　为图 9 - 8 所示的街道图找出最短的邮递路线.假定邮局位于点 A.

解　最短的邮递路线不存在(不存在的原因后面给出).

看来不是在任何情况下都能找到理想的邮递路线的.那么在什么情况下才能找到理想的路线呢?这使我们联想到一个有名的数学游戏,叫作一笔画问题.

9.1.2　一笔画问题

一笔画问题是,什么样的图形可以一笔画成?要求是笔不离纸,而且每条线只画一次,不准重复.

看图 9－9 的三个图.读者通过试验会断定,"田"和"品"不可能一笔画成,而"串"却可以.理想的邮递路线问题比一笔画的要求还多一点,就是最后要回到起点.以后我们会看到,这个差别不是关键的.这样,一笔画问题可以视为最短邮递路线的一部分.我们把最短的邮递路线问题暂时放在一边,先花较多的篇幅讨论一笔画问题,然后再回来分析最短的邮递路线问题.一笔画问题是欧拉提出并解决的.这涉及到哥尼斯堡七桥问题.

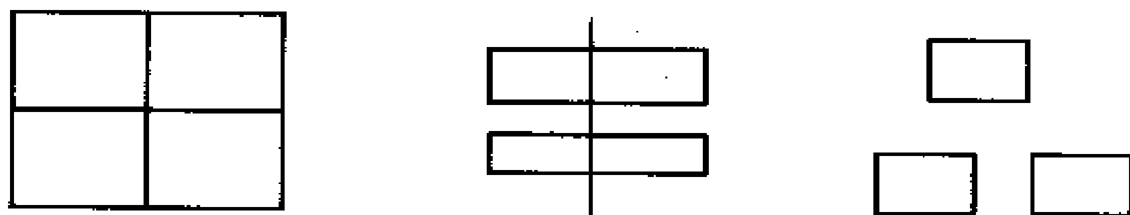


图 9－9

9.1.3 哥尼斯堡七桥问题

故事发生在 18 世纪的哥尼斯堡城,现在是俄罗斯的加里宁格勒市.这座城市建立的在普雷格尔河畔,由 4 块分开的土地构成,中间有七座桥相连,如图 9－10 所示.当时那里的居民热衷于一个难题:一个散步者怎样才能一次走遍七座桥,每座桥只走过一次,最后回到出发点?这个问题看起来不难,谁都愿意试一试,但是谁也找不出答案.

问题最后由欧拉解决.千百人的失败,使他猜想,也许那样的走法根本不存在.1736 年,他证明了他的猜想,并在圣彼得堡科学院作了一次报告.

他用点 A 表示岛,点 B 表示河的左岸,点 C 表示河的右岸,D 表示两条支流间的区域,如图 9－10 所示.用联结两点的线来表示联结两块陆地的桥,由此得到一个由七条线组成的图形(图 9－11).这

样，哥尼斯堡七桥问题就变成了一个一笔画问题：能不能一笔画出这个图形，并且能最后返回起点？

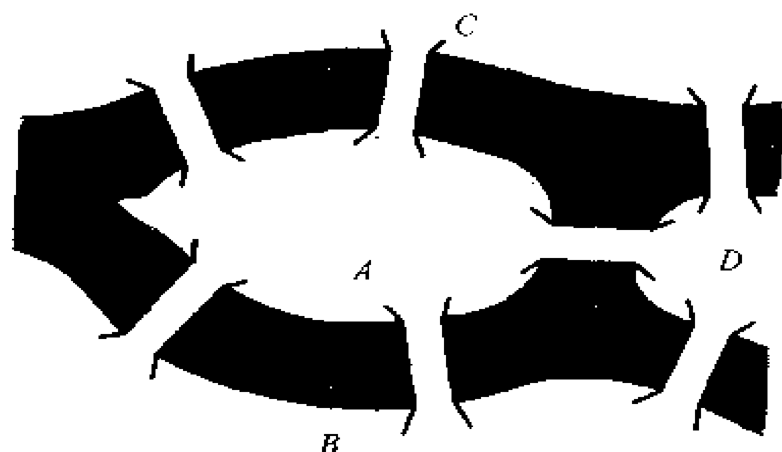


图 9－10

现在我们分析一下画图的过程．如果我们从某一点出发一笔画出了一个图形，而到某一点停止，那么中间每经过一点，总有进去的一条线和出来的一条线．所以除了起点和终点这两点外，这个图形上的每一点都应该和偶数条线相联结．如果起点和终点相重合，那么起终点也应该和偶数条线相联结．如果起点和终点不重合，那么起终点将与奇数条线相联结．

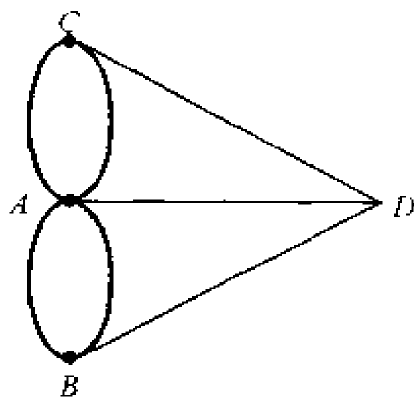


图 9－11

但是图 9－11 上的三个点 B、C、D 都和三条线相联结，而点 A 和五条线相联结，都是奇数条线．这当然不可能一笔画出，并回到起点．即使不回到起点，也不可能一笔画出．

正是作了这样的分析之后，欧拉才断定，不管要求不要求回到起点，不重复地一次走遍那七座桥总是不可能的．

七桥问题,或者一笔画问题,明显地是一个几何问题.但是这种几何问题却是欧几里得几何所没有研究过的.因为欧氏几何中研究的图形都由直线和圆组成,讨论的是长度和角度等性质.在七桥问题里,桥的准确位置和长度是无关紧要的.要紧的是每两块陆地间有几座桥.一笔画问题里线段的长短曲直也无关紧要,要紧的只是有几个分岔点,两点间有几条线相联结.也可以说,要紧的只是点线间的相互位置,或相互联结的情况.所以欧拉把这类几何问题的研究叫作位置几何学.对于这么一类新鲜的几何问题,欧拉当然不满足于只解答一个七桥问题.他继续钻研,终于找到了一个简便的原则,可以鉴别任一图形能不能一笔画出.这就是一笔画定理.在讲一笔画定理前,先引进一些术语和记号.

9.1.4　网络

我们讨论的图形都是由线段构成的.对这类对象我们有一个专门的名词来称呼它,叫作网络.为了明确起见,我们给它下一个定义.

定义　网络是由有限条线段组成的图形,每一条线段都有两个不同的端点.这些线段叫作网络的弧,它们的端点叫作网络的顶点.

例　图 9 – 12 就是有 8 条弧和 5 个顶点的网络.

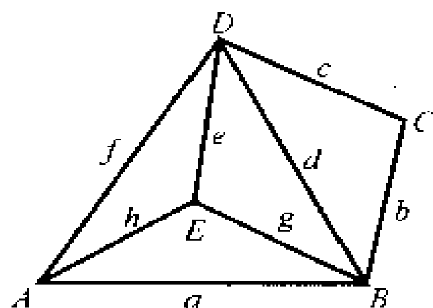


图 9 – 12

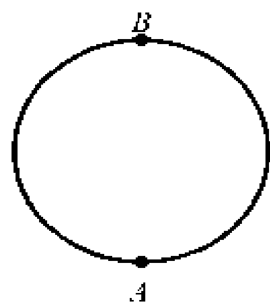


图 9 – 13

欧氏几何中许多由线段构成的图形,在适当分解成弧以后,都能

看成网络.网络的弧不能没有端点.例如整个的圆周就没有端点,它不是一个网络.但我们可以把它分解为有两个端点和两个弧的网络,如图9-13所示.网络的弧也不能只有一个端点,如两端重叠的闭弧不包括在其内.网络的弧必须有两个端点.

为什么要这样抽象地提出网络的概念呢?这是因为有了网络的概念,不但使我们以后说起话来方便;而且可以扩大我们讨论的范围.按照这个定义,不单纸上的图可以看成网络,电子仪器的复杂电路也可以看成网络,一个国家的铁路网、公路网、水运网,一个区的街道网,都可以看成网络.不但有平面上的网络,也有空间中的网络.所以我们对一般网络进行研究所得到的种种结论,就可以应用到所有这些具体的网络.

为叙述方便起见,我们引入下面的记号.网络中的顶点用大写字母表示,弧用小写字母表示.$l = AB$ 表示以 A、B 为端点的弧.当然,若 $l = AB$,那么也有 $l = BA$.但是要注意,$l_1 = AB$,$l_2 = AB$,未必有 $l_1 = l_2$,因为以 A,B 为端点的弧可以不止一条,如图9-11中以 A,B 为端点的弧有两条.网络本身用黑体大写字母表示.

我们把互相衔接的一串弧叫作一条路.写得更精确些,网络中的一条路可以写成$(l_1, l_2, \cdots, l_k)$的形式,其中的弧两两不相同,并且每条弧 $l_i (i = 2, \cdots, k)$ 都与 l_{i-1} 在点 A_{i-1} 相接,另一端 A_i 与 l_{i+1} 相接.路也用黑体大写字母表示,记作 $\mathbf{W} = (l_1, l_2, \cdots, l_k)$,$l_i = A_{i-1}A_i (i = 1, 2, \cdots, k)$.$A_0$ 和 A_k 叫作这条路的两端,或者说,这条路连接 A_0 和 A_k.若 $A_0 = A_k$,则这条路叫作闭路.若顶点 $A_0, A_1, \cdots, A_{k-1}$ 又互不相同,那么这条闭路就叫作一个圈.

例 图9-12中,(a, d, e, g)是路;(a, d, g, e)不是路,因为 g 不是接在 d 的终点处;$(a, g, h, f, c, b, d, e, h)$不是路,更不是闭路,因为弧 h 出现了两次;(a, b, c, e, g, d, f)是闭路而不是圈,因为中间的顶点 B,D 出现两次;(a, d, e, h)是一个圈.

回到一笔画问题.现在不难看出,一笔画问题相当于,给定了一

个网络,问有没有可能把所有的弧排成一条路.要求一笔画成后回到起点,相当于要求把全部弧排成一条闭路.这就把一笔画问题用网络语言表达出来了.同时也把问题推广了.原来的一笔画问题只是对平面上的图形来说的,而现在的提法对任何网络都有意义,不必限于平面上的网络.

定义 如果一个网络的全部弧可以排成一条路,那么这个网络就叫作一个一笔画.

注 网络中的路不必是闭路.

有些图形,像前面图中的"品"字,所以不能一笔画成,显然是因为它不是连成一片的.可见能一笔画成的图形必须是连成一片的.所谓连成一片到底是什么意思?我们来下个定义.

定义 一个网络称为连通的.如果它的任意两个顶点都可用一条路连接起来;否则称为不连通的.

例如,全国大陆的铁路形成的网络是连通的.但全国铁路形成的网络不是连通的,因为台湾省和海南省的铁路不与大陆的铁路相连.

如果某网络是由几个网络并成的,而这几个网络彼此没有公共的顶点和弧,那么这网络就一定不是连通的.不连通的网络总是由几个互不相交的连通网络并成的.这几个连通的网络叫作原网络的分支.例如,全国铁路形成的网络至少由三个分支并成.

我们注意到,一笔画问题又与顶点处的弧的个数有关,即与顶点处的分岔情况有关.

定义 以某个顶点为端点的弧的个数叫作该顶点的叉数.

在图 9 - 12 中,B 是 4 叉顶点,C 是 2 叉顶点,E 是 3 叉顶点.顶点的叉数又有奇偶之分.叉数是奇数的顶点叫作奇顶点,叉数是偶数的顶点叫作偶顶点.奇顶点的多少和一笔画问题有极大的关系.没有奇顶点的网络叫作偶网络.

9.1.5 一笔画定理

有了上面的准备后我们可以讲述一笔画定理了.我们分成两个

问题讲：

1）具有哪些特征的网络不是一笔画？

2）具有哪些特征的网络是一笔画？

1）网络不是一笔画的特征　要证明一个网络不是一笔画，就是要证明在这个网络中全部的弧不可能排成一条路，或者说网络中不存在一条路包括所有的弧．这就是说，不能只承认你自己没有找到这样的路，而且还要断定别人也不会找到这样的路．找不到的原因不是主观能力不够，而是客观上不存在．这种"不存在性"定理在数学中是很多的．前面几讲中讲的几何三大难题、费马大定理都属于这一类定理．证明"不存在性"定理的最常用的方法是反推法：先证明如果"存在"，则必须如何如何；然后说，现在并不如何如何，所以"不存在"．前面证明七桥问题不可能有解，用的就是这个办法．

这样一来，我们的第一个问题可以换个问法：一笔画必须具备哪些性质？下面的定理就是回答这个问题的．七桥问题的解决实际上用的正是这个定理．

定理 1　一笔画必须是连通的，并且奇顶点的个数是 0 或 2．

证　a）连通性．按定义，一笔画的全部弧可以排成一条路

$$Z = (l_1, l_2, \cdots, l_k), l_i = A_{i-1}A_i, i = 1,2,\cdots,k.$$

这里 $A_0, A_1, \cdots, A_k$ 包括网络的全部顶点（可能有重复），因为任何一个顶点至少是一条弧的端点．任取两点 $A_i, A_j, i < j$ 它们可用路 $(l_{i+1}, \cdots, l_j)$ 联结起来．所以网络是连通的．

b）奇顶点的个数是 0 或 2．从网络中任取一个顶点 A．可能出现两种情况：A 是内点；A 是起点或 A 是终点．若 A 是内点，则 A 一定是偶顶点．因为有一条线段进入 A，必有一条线段离开 A．所以内点都是偶顶点．

现在设 A 是起点或 A 是终点．这时又有两种情况：A 既是起点又是终点；A 只是起点或只是终点．在第一种情况下，与 A 是内点的情况一样，它是偶顶点．这时网络没有奇顶点，或只有 0 个奇顶点．

在第二种情况下，若 A 只是起点，则从 A 出去的线段比从 A 进去的线段多一条，从而 A 是奇顶点．若 A 只是终点，则从 A 出去的线段比从 A 进去的线段少一条，从而 A 也是奇顶点．这时网络有两个奇顶点．定理证毕．

定理 1 给出了一笔画的必要条件．有了这个定理我们就可以断定，不连通的，或者奇顶点的个数不是 0 或 2 的网络一定不是一笔画．

例　"品"字不是连通网络，所以不是一笔画．"田"字中有 4 个奇顶点，不是一笔画．七桥问题中有 4 个奇顶点，也不是一笔画．

那么，一个网络是连通的，它的奇顶点数又恰是 0 或 2，它就一定是一笔画吗？欧拉断定：是的．这就回答了我们的第二个问题：根据哪些特征可以断定一个网络是一笔画．怎么证明呢？要证明一个网络是一笔画，就应该证明这个网络中的全部弧可以排成一条路，或者说，这个网络中存在一条路包括全部弧．

这种"存在性"定理在数学中是很多的．前面我们曾经证明过无理数的存在性，超越数的存在性，等等．这些定理，有的很容易证明，有的却非常困难．在代数部分，我们还将证明代数基本定理，那也是一个很难的存在性定理．"存在性"定理所肯定的只是"存在"，它不回答存在多少个的问题．

2）网络是一笔画的特征　　证明"存在性"定理，最简单的办法是直接找出一个来．前面图 9－3 的一笔画存在性就是这样证明的．但是用这种办法证明欧拉的判断是有困难的．对于一个具体的网络，我们可以设法找出包含所有弧的路，以此证明它是一笔画．现在要证明的是具有某种性质的网络都是一笔画，而有这种性质的网络多得不可胜数，怎么可能一一找出它们的路呢？欧拉的判断的妙处正在于，使能够在碰到某些很复杂的网络时，不用找路就能很快地断定它是不是一笔画．

我们把欧拉的判断分成两个定理来讲．证明的线索是，间接地指

出一种把全部弧排成路的办法.在推理的过程中,网络中弧个数的有限性起着重要的作用.

定理2 若网络 G 是连通的偶网络,则 G 的全部弧可以排成一条闭路.

证 分三步来证.

(1) 先证:若 A_0 是 G 中任一顶点,则在 G 中一定能找到一条从 A_0 到 A_0 的闭路.

事实上,任取第一条弧 $l_1 = A_0A_1$,再取第二条 $l_2 = A_1A_2$,如果需要的话再取第三条 $l_3 = A_2A_3$,等等.这样就作出一条路

$$Z = (l_1, l_2, \cdots, l_k), \quad l_i = A_{i-1}A_i \quad (i = 1, 2, \cdots, k).$$

当 $A_k \neq A_0$ 时,这条路一定还可以延长;因为这时 Z 中以 A_k 为端点的弧是奇数条,而 A_k 是偶顶点.所以一定还可以找到不在 Z 中的弧 $l_{k+1} = A_kA_{k+1}$,而得到$(l_1, \cdots, l_k, l_{k+1})$.这条路越作越长,但 G 中总共只有有限条弧,所以总有一个时候作不下去了.设这时的路是

$$(l_1, l_2, \cdots, l_m), \quad l_i = A_{i-1}A_i \quad (i = 1, 2, \cdots, m),$$

那么 A_m 一定与A_0 重合,即这条路是闭路.

(2) 再证:若 G 是连通的偶网络,Z 是 G 中的一条闭路,它没有包含所有的弧,则一定能找到一条闭路 H,它含有的弧比 Z 含有的弧多.

设 $Z = (l_1, l_2, \cdots, l_k)$, $l_i = A_{i-1}A_i$, $A_k = A_0$.从网络 G 中把Z 的弧抹掉,抹掉后剩下的网络叫 G'. G' 也一定没有奇顶点,即它也是偶网络.为什么呢?因为从 G' 中任取一个顶点A,在 G 中有偶数条弧以 A 为端点;Z 是闭路,所以 Z 中也有偶数条,或者 0 条弧以 A 为端点,于是 G' 中以A 为端点的弧也是偶数条.

Z 和G' 一定有公共点,否则 G 由两个没有公共顶点的网络Z 和 G' 并成,它就不会连通了,这和假设矛盾.设 $B_0 = A_r$ 是一个公共点,$1 \leqslant r \leqslant k$.

根据(1),在 G' 中一定有一条从 B_0 到 B_0 的闭路

$$Z' = (l'_1, \cdots, l'_s), \quad l'_j = B_{j-1}B_j, j = 1, \cdots, s, B_s = B_0.$$

于是

$$\overline{Z} = (l_1, \cdots, l_r, l'_1, \cdots, l'_s, l_{r+1}, \cdots, l_k)$$

就是 G 中的一条闭路,它包含的弧比 Z 多.

(3) 定理的证明.根据(1),在 G 中可以取到一条闭路 Z.若 Z 包含了 G 中的全部弧,则定理已经被证明.若 Z 没有包含 G 中的全部弧,则根据(2),在 G 中可以取一条比 Z 含更多弧的闭路 Z_1,若 Z_1 还没有包含 G 的所有弧,再如法取 Z_2,这样继续扩大下去.由于在 G 中总共只有有限条弧,总有一个时候扩大不下去了.这时得到的闭路必然包含 G 中的所有弧.证毕.

定理 3 如果一个连通网络 G 的奇顶点数是 2,那么这个网络一定是一笔画.

证 这个定理的证明可以仿照定理 2,证明分为三步.(1) 先证:若 G 是有两个奇顶点 A 和 B 的网络,那么在 G 中一定能找到一条从 A 到 B 的路.(2) 再证:若 G 是有两个奇顶点 A 和 B 的连通网络,W 是从 A 到 B 的一条路,它没有包含 G 的全部弧,那么一定还能找到一条从 A 到 B 的路 $\overline{W}$,它包含的弧比 W 的弧多.(3) 定理本身的证明.作为练习,建议读者自己去证明.

还有一种巧妙的证法,叙述如下.办法是在网络中加一条联结奇顶点 A 和 B 的弧 l.(注意,如果我们只考虑平面上的网络,限定在平面上找这样的弧,又要求它和网络中的其它弧不相交,那么这样的弧可能找不到,图 9-14 就是一个例子.现在我们在空间加这条弧,这总办得到.这是不局限于讨论平面网络的一个好处.)添加 l 后所得的网络就是一个没有奇顶点的连通网络了.根据定理 2,它的所有弧可以排成一条闭路.不妨把

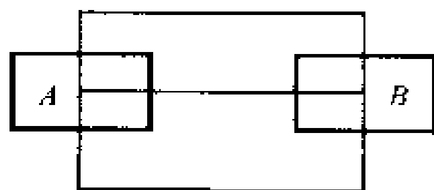


图 9-14

l 排在前面：$(l, l_1, \cdots, l_k)$. 于是 G 的全部弧可以排成一条路 $(l_1, \cdots, l_k)$. 证毕.

定理 1 告诉我们，一笔画必须是连通的，而且奇顶点的个数是 0 或 2. 定理 2,3 又告诉我们，由这两个性质就足以断定一个网络是一笔画. 合起来就是

一笔画定理 一个网络是一笔画的充要条件是：它是连通的并且奇顶点的个数是 0 或 2.

一笔画定理把一笔画问题彻底、漂亮地解决了. 彻底指的是，它给出了充分必要条件；从而把一笔画和非一笔画的界限彻底划清了. 漂亮指的是，它给出的充分必要条件简单明了，很容易检验，用起来很方便. 即使一个有几十条弧的网络，要判别它是不是一笔画，也花不了一二分钟.

从一笔画的理论中容易看出，在邮递路线问题中，可能找到理想的不重复的路线的充分和必要的条件是，投递范围的街道网连通且没有奇顶点.

9.1.6 多笔画

我们可以进一步提出问题：如果一个连通网络的奇顶点数不是 0 或 2，那么一笔是画不成了. 要多少笔才能画成呢？看来笔数一定和奇顶点的个数有关. 我们先证明一个关于奇顶点个数的定理.

定理 4 一个网络中的奇顶点的个数一定是偶数.

证 设网络中弧的个数是 a，叉数是 m 的顶点的个数是 b_m. 我们来数一数网络中顶点的个数. 一方面，每条弧有两个端点，所以共有 $2a$ 个端点. 另一方面，m 叉顶点是 m 个弧的端点. 所以

$$2a = b_1 + 2b_2 + 3b_3 + 4b_4 + 5b_5 + 6b_6 + \cdots$$

（注意，和的项数不是无穷，而是有穷）这是一个偶数. 从这个偶数中减去偶数

$$2b_2 + 2b_3 + 4b_4 + 4b_5 + 6b_6 + 6b_7 + \cdots$$

得到的数

$$b_1 + b_3 + b_5 + \cdots$$

还应该是偶数，而这个数恰是网络的奇顶点的个数. 证毕.

定理 5 设 G 是一个有 $2n$ 个奇顶点的连通网络，$n > 1$. 那么，G 中的全部弧可以排成 n 条路，而且至少 n 条路.

证 把 $2n$ 个奇顶点分成 n 对：$A_1, B_1, A_2, B_2, \cdots, A_n, B_n$. 给 G 添上 n 条新的弧 $l_i^0 = A_iB_i$. 这样一来，得到一个新网络 G^*，它是一个没有奇顶点的连通网络. 把 G^* 的全部弧排成一条闭路 Z，然后从中把那 n 条新添的弧去掉. 于是闭路 Z 被切成 n 段. 每一段是 G 中的一条路，这 n 条路包含了 G 的全部弧.

至少要 n 条路也不难证. 因为如果能排成 q 条路，而每条路最多有两个奇顶点，那么 G 中的奇顶点的个数将 $\leqslant 2q$. 所以 $2n \leqslant 2q$，即 $q \geqslant n$. 证毕.

9.1.7 偶网络

一笔画和偶网络有联系也有区别. 一个一笔画不一定是偶网络，因为一笔画可以有两个奇顶点，而偶网络没有奇顶点. 另一方面，一笔画必须是连通的网络，偶网络可以是不连通的. 现在我们讲一个关于偶网络的定理，后面讲邮递路线问题时要用到.

定理 6 网络 G 是偶网络的充要条件是，G 的所有弧可以排成若干个圈，这些圈彼此没有公共弧.

证 条件的充分性是容易看出的. 即，如果 G 的所有弧可以排成若干个圈，这些圈彼此没有公共弧，那么 G 没有奇顶点，因而是偶网络.

现在证必要性. 即如果 G 没有奇顶点，则 G 的所有弧可以排成若干个圈，这些圈彼此没有公共弧. 证明分两步.

(1) 先证：在偶网络 G 中任取一个顶点 A_0，一定能找到从 A_0 到 A_0 的圈.

根据定理 2 证明中的(1),G 中有 A_0 到 A_0 的闭路.取出这种闭路中弧数最少的一条(或弧数最少的几条之一)

$$Z = (l_1, l_2, \cdots, l_k), \quad l_i = A_{i-1}A_i, A_k = A_0.$$

这条闭路一定是圈.因为否则会有 i,j 两数,$1 \leqslant i < j \leqslant k$,使 $A_i = A_j$.从而

$$(l_1, \cdots, l_i, l_{j+1}, \cdots, l_k)$$

是比 Z 短的 A_0 到 A_0 的闭路,这与 Z 的最短性矛盾.

(2) 定理的证明.在 G 中取一个圈 Z.如果 Z 已包含 G 的全部弧,则定理已成立.否则,从 G 中把 Z 抹掉,剩下的网络 G_1 还是没有奇顶点的网络.再从 G_1 中取一个圈 Z_1,如果它包含 G_1 的全部弧,那么 G 的全部弧可以排成两个圈 Z 和 Z_1.否则从 G_1 中把 Z_1 抹掉,又得一个网络 G_2.这样作下去,引出一串偶网络

$$G, G_1, G_2, \cdots.$$

由于弧的个数一个比一个少,而 G 中只有有限条弧,所以这串网络总有个尽头 G_r.在 G_r 中取出的圈 Z_r 一定包含 G_r 的全部弧.于是 G 中的全部弧可以排成 $r+1$ 个圈 $Z, Z_1, \cdots, Z_r$.证毕.

9.1.8 再论邮递路线问题

我们已经比较深入地讨论了一笔画问题,现在回到最短邮递路线问题,讲奇偶点图上作业法.最短邮递路线问题用网络的语言可改说为:

给定一个连通的网络(投递范围的街道图),每条弧有个长度(在一笔画理论中,从来没有提到过弧的长度,因为在那里弧的长度是不起作用的),要求从某一顶点出发走过网络的所有弧(容许重复,这是和一笔画不同的地方),并且最后返回起点,问怎样走法才能使走的路程最短?

根据一笔画理论,我们可以这样来分析问题.

如果这网络中没有奇顶点,那么根据定理 2,它可以从任一点开

始没有重复地一笔画出,最后回到起点.前面已经说过,这样理想的路线是最短的.

如果网络中有奇顶点,如图 9－15 所示(图中有 4 个奇顶点),那么要从一点开始走遍各弧回到起点,必须有重复.走的路线长短,就看路线中重复部分的长短.如果把某一种走法的重复路线添在图上,如图 9－16 所示,那么所有的顶点一定都变成偶顶点了.反过来说,若先在网络中的某些弧上添一些"重复"弧,使得添加后的网络中所有的顶点都是偶顶点,那么根据定理 2,添加后的网络就能够不重复地一次走遍并返回起点,也就是代表一种走法,如图 9－17 所示.

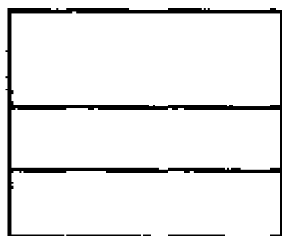


图 9－15

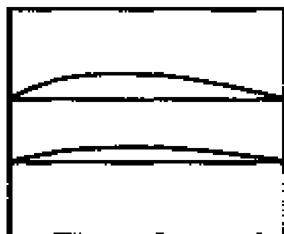


图 9－16

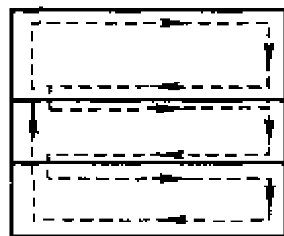


图 9－17

这样一来,利用一笔画的理论,问题归结为,一个连通网络有 $2n$ 个奇顶点,要在某些弧上添一些重复弧(每条弧上添的重复弧条数不限,且重复弧与原弧同长),使得添弧后的网络没有奇顶点.问怎样添法能使重复弧的长度最短?下面就来讨论.

9.1.9 奇偶点网上作业法

为叙述方便计,我们把添重复弧后没有奇顶点的添法叫作一个解,重复弧长度最短的解叫作最优解.先给一个命题.

命题 1 解总是存在的.

解的存在性在直觉上是明显的.因为只要容许重复,总可以走遍各街道后返回邮局.作为数学证明,我们具体指出一种作解的方法.把原网络中的 $2n$ 个奇顶点随便分成 n 对:$A_1,B_1;A_2,B_2;\cdots;A_n$,

B_n. 取 n 条路 $\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_n$ 分别连接. $A_1, B_1; A_2, B_2; \cdots; A_n, B_n$. 由于网络是连通的,这些路总是存在的. 把这些路中的弧添到原网络中去,就是一个解.

由此可见,具体找一个解是很容易的. 问题只是如何求最优解. 读者或许会这样想,只要每条 $\boldsymbol{W}_i$ 都是连接 A_i, B_i 的最短的路,不就得到最优解了吗?其实问题不那么简单. 把 $2n$ 个奇顶点重新配对,或许会得到更短的路.

例 图 9 - 18 是一个街道图,图上注明的数字是街道的长短米数. 这个网络中有 8 个奇顶点. 我们把它们像图 9 - 19 那样对奇点进行分类,添上连接 A_i, B_i 的最短路 $\boldsymbol{W}_i$($i = 1, 2, 3, 4$) 而得出解(1). 显然它不是最优解.

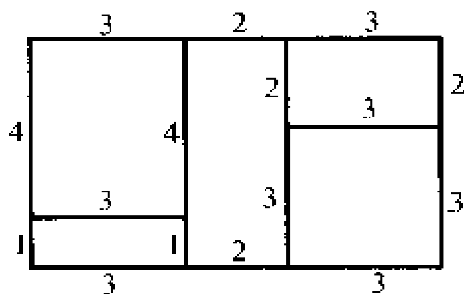


图 9 - 18

怎样求得最优解呢?我们知道,无论做什么事情,写文章也好,做习题也好,不会一下子就十全十美的,总有一个琢磨推敲,修正缺点逐步完善的过程. 同样道理,一个比较好的解常常是从修改一个比较差的解而得到的. 所以我们这样做,随便拿一个解来,逐步修改它,使它缩短.

我们看图 9 - 19 上的解(1). 这里 B_3 和 A_4 之间有两条重复弧,这就是可以修改的地方. 由此得到下面的命题.

命题 2 如果一个解有重叠的重复弧,即在某条弧上有多于一条的重复弧,则这个解可以改进,从而这个解不是最优解.

证 证明很简单. 如果这个解在某条弧上至少有两条重复的弧,那么去掉其中的一对,还是一个解. 因为去掉一对弧,不会改变网络的奇偶性,即网络仍是偶网络. 去掉一对弧后,重复弧的总长度显然是缩短了.

例 图 9 - 19 中,解(2) 就是从解(1) 中去掉一对重复弧 B_3A_4

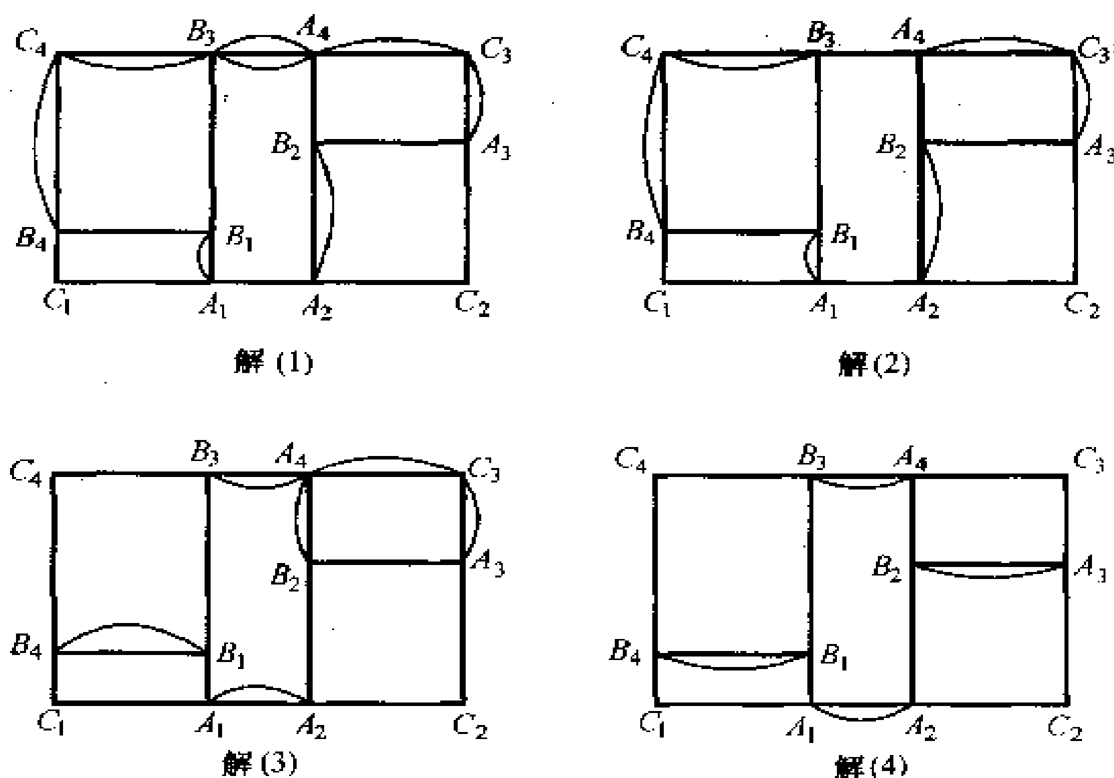


图 9 - 19

后得到的.

这个浅显的命题却有一个重要的推论.

推论 最优解是存在的.

注 这件事需要证明.因为并非任何东西都有个"最".例如最大的整数不存在;最小的正数也不存在.

证 根据命题2,有重叠重复弧的解一定可以改善,使重叠重复弧的条数减少.但是,在一个解里面,重叠重复弧的条数总是有限个,所以经过有限多次修改后,总会改成一个没有重叠重复弧的解.

另外,由于原网络中只有有限条弧,没有重叠重复弧的解只能有有限个.所以其中必有一个(或几个,但至少有一个)解,它的重复弧的总长度在这有限个解中是最短的.这个解就是最优解.

注 又是有限性在这个证明中起了关键的作用.在有限个解中找最好的,是一定找得到的.

总之,命题 2 告诉我们两点:1) 最优解中没有重叠的重复弧;2) 最优解存在.

找最优解只需在没有重叠重复弧的解里面去找.但是怎样去找呢?下面提供一个原则.

命题 3 设有一个解,它没有重叠的重复弧,并且在原网络的某个圈上,重复弧长度的和超过圈长的一半,那么这个解可以改善,即它一定不是最优解.

证 在命题指出的那个圈上,没有重复弧的弧的长度的和一定小于圈长的一半.既然这样,我们就可以在这个圈上把原来的重复弧去掉,给原来没有重复弧的弧添上一条重复弧.这样修改后得到的仍是一个解,并且没有重叠的重复弧,而重复弧的长度的总和却减少了.为什么仍是一个解呢?这只需说明,这种修改不影响顶点的奇偶性.若某个顶点是两个有重复弧的弧的交点,那么去掉重复弧后,这个顶点的叉数减 2.这当然不会改变它的奇偶性.若另一顶点是一个有重复弧的弧与一个没有重复弧的弧的交点,那么修改后这个顶点的叉数不变,因为一边减去 1,而另一边加上 1.证毕.

例 看图 9－19 的解(2) 中的圈 $A_1A_2B_2A_4B_3C_4B_4B_1A_1$.这个圈的总长度是

$$2+3+2+2+3+4+3+1=20.$$

重复弧的长度的和是 $3+3+4+1=11$,它大于总长度的一半.按命题 3 的证明中提供的办法修改后得解(3),它比解(2) 好.

例 看图 9－19 的解(3) 中的圈 $A_3C_3A_4B_2A_3$.这个圈的总长度是

$$2+3+2+3=10.$$

重复弧的长度的和是

$$2+3+2=7,$$

它大于总长度的一半.按命题3的证明中提供的办法修改后得解(4).

从任意一个解出发,按命题2去掉重叠弧,再反复利用命题3进行修改.由于没有重叠重复弧的解只有有限多个,不可能无限地修改下去,所以经过若干次修改后,一定能得出一个解,它既没有重叠的弧,在每个圈上的重复弧长又都不超过圈长的一半.图9-19的解(4)就具有这个性质.这样的解能不能断定它就是最优的呢?当然,利用刚才的两种办法是不能把它再改善了.可是我们不知道还有没有别的办法可以改善它.下面的两个定理指出,这样的解确实不能再改善了.或者说,这样的解就是最优解.

定理7 如果两个解都满足下面两个条件:

(1) 没有重叠的重复弧;

(2) 在原网络的每个圈上,重复弧的长度和不超过圈长的一半,那么,这两个解中的重复弧的长度的总和相等.

证 为方便计,把这两个解记为 a、b,原网络记为 G.我们要证明:

解 a 的重复弧的总长 = 解 b 的重复弧的总长.

首先,我们把解 a 和解 b 里所有的弧并在一起.这时所有原来的弧都成了重叠的重复弧.把它们都去掉.其结果是只剩下 a,b 的重复弧,这些弧构成一个新网络,它是 G 的一部分,并仍是一个偶网络.

图9-20是个例子.(1)是街道图,即原网络 G.(2)和(3)是合于上面条件的解 a 和解 b.(4)是新网络.

其次,在新网络中去掉重叠的重复弧,不重叠的重复弧组成一个网络 G',如图(5)所示.解 a 和解 b 的重复弧的总长度的差,就等于解 a 和解 b 在 G' 上的重复弧的长度的和的差.

从 G' 中任取一个圈,那么这个圈的每一个弧都是解 a 或解 b 的重复弧.所以在这个圈上,a 的重复弧的长度和加上 b 的重复弧的长

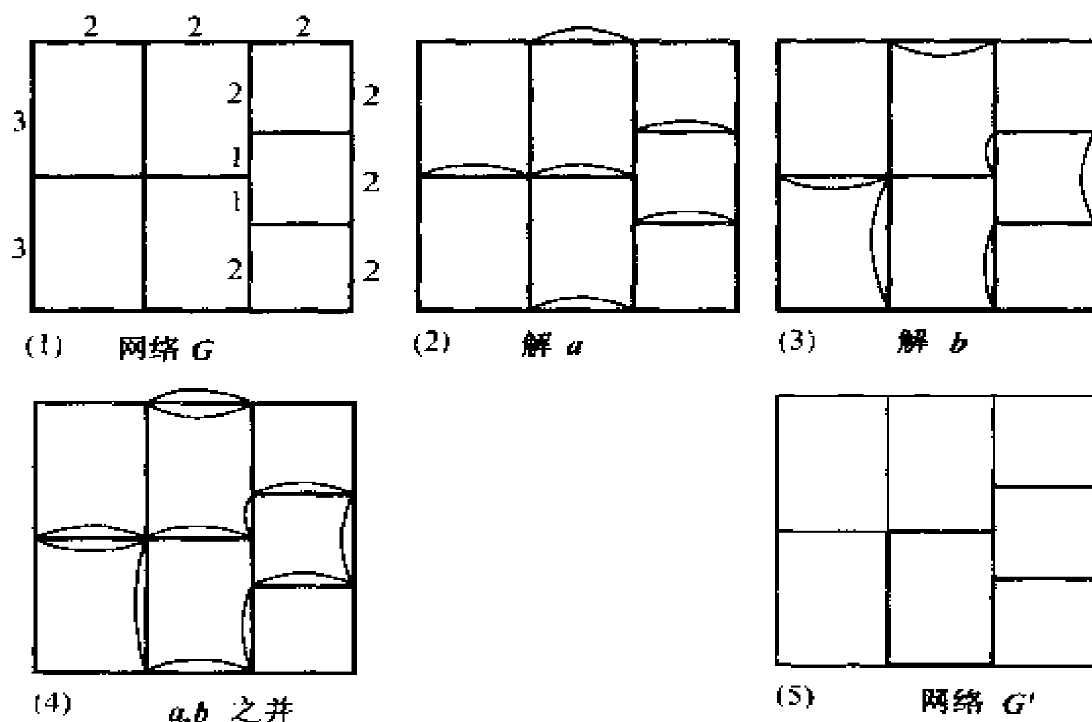


图 9 - 20

度和等于圈长.但 a 的重复弧的长度和与 b 的重复弧的长度和都不超过圈长的一半,因此在这个圈上,a 的重复弧的长度和与 b 的重复弧的长度和恰好相等.

注意,G' 是没有奇顶点的网络.这是因为,任取一个顶点 A,解 a 和解 b 中以 A 为顶点的重复弧的条数或者同是奇数,或者同是偶数,决定于 A 是原网络的奇顶点还是偶顶点.所以,并起来一共是偶数条.在 G' 中以 A 为端点的弧就是从这偶数条里减去几对相互重叠的弧(这也一定是偶数条),所以以 A 为端点的弧是偶数条.这说明 G' 是偶网络.

根据定理 6,G' 的弧可以排成若干个圈.上面已指出,G' 的每个圈上 a 的重复弧的长度和与 b 的重复弧的长度和相等,所以在整个 G' 上,a 的重复弧的长度和与 b 的重复弧的长度和也恰好相等.由此

可见,a 的重复弧的长度总和与 b 的重复弧的长度总和相等.证毕.

定理 8 一个解是最优解的充要条件是,它满足定理 7 的条件(1) 和(2).

证明 必要性.命题 2 和命题 3 指出,最优解必须满足条件(1)和(2).

充分性.我们来证明满足条件(1) 和(2) 的解必是最优解.

由命题 2 的推论,最优解是存在的.譬如说 a 是一个最优解.根据已证的必要性部分,a 一定满足(1) 和(2).今设 b 是满足(1) 和(2) 的另一解.根据定理 7,a 的重复弧的长度总和与 b 的重复弧的长度总和相等,所以 b 一定也是最优解.证毕.

到此为止,可以说最短邮递路线的问题初步解决了.我们证明了最优解的存在性,并且具体给出了一种求解并将解改善的方法,证明了按这种方法修改有限次后一定能够得到最优解.归纳起来,寻找最好邮递路线的步骤如下:

1) 画出邮递范围的街道图;

2) 找出街道图的奇顶点;

3) 添上重复弧,把奇顶点对对相联;

4) 按命题 2 去掉重叠的重复弧;

5) 按命题 3 反复修改,直到不能修改;

6) 将所得的有重复弧的网络一笔画出.

这种方法叫作奇偶点作业法.

为什么只说初步解呢?因为这种找法还不够理想,用起来不够方便.原因是步骤 5) 可能相当长,或者说定理 7 的条件(2) 不容易检验.稍为复杂一点的图检查起来就很困难,因为其中的圈可以多到几百个.从这个角度看,奇偶点图上作业法解决最短邮递路线问题不及一笔画定理解决一笔画问题那样好.

9.1.10 什么是拓扑学

一笔画问题来自拓扑学.这个问题与我们以往碰到的欧氏几何

的问题有明显的不同.在欧氏几何中我们要考虑长度和角度等度量性质.但是一笔画问题与长度和角度无关,它只涉及到网络的布局.利用一笔画问题的研究,我们解决了最短邮递路线问题.在邮递路线问题中,我们考虑了长度,所以它不再是一个纯拓扑的问题了.

那么,什么是拓扑学呢?

为了说明拓扑学是研究什么的,先回顾一一对应的定义.设 A,B 是两个点集.称映射 $f:A\to B$ 是一一对应的,如果集合 B 中每一个点恰好是集合 A 中一点的象.即,集合 A 中两个不同的点映射为集合 B 中两个不同的点;且集合 B 中的每一个点都是集合 A 中某一点的象.对于一个一一对应的映射 $f:A\to B$ 可以定义逆映射 $f^{-1}:B\to A$.

定义　称映射 $f:A\to B$ 是同胚映射,如果它既是一一对应的,又是双方连续的.

直观地说,同胚可以看作从一个集合到另一个集合的这样的映射,它既不断开也不重叠.对二维的集合我们可以把它想象为橡皮膜作成的.可以用任意方式压缩个和拉伸,只是不能使它断裂,也不能把不同的点"粘合"成一点.如果在这些条件下,能使图形 A 和图形 B"重合",则它们同胚.

例　图 9-21 的 4 个 1 维图形是同胚的.



图 9-21

例　图 9-22 的 4 个 2 维图形是同胚的.

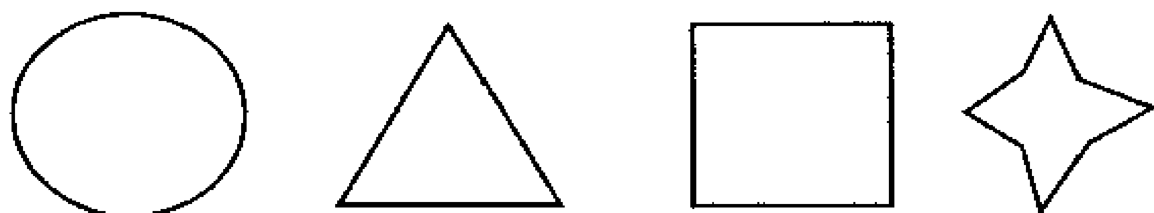


图 9 – 22

例 球面,立方体表面,圆柱体表面都是同胚的,但它们不同胚于环面.环面与壶铃的表面同胚(图 9 – 23).

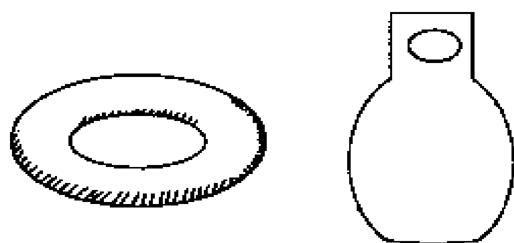


图 9 – 23

拓扑学作为一门学科形成于 19 世纪末,主要是由法国大数学家庞加莱开创的.但拓扑学的起源可追溯到欧拉和黎曼.庞加莱曾经这样来确定拓扑学的内容:"拓扑学是一门科学,它不仅使我们认识通常空间中几何图形的定性性质,而且也能使我们去认识高于 3 维空间几何图形的定性性质.在 3 维空间中,拓扑学几乎是直观的.反之,对于高于 3 维的空间,拓扑学就显得难以琢磨了."

我们简单地比较一下拓扑学中的同胚概念与欧氏几何中的全等概念.在欧氏几何中,讨论等距映射,即保持两点间距离不变的映射,并把它叫作运动.每个图形运动的结果是作为整体不改变距离而变到新的位置.两个图形是全等的,指的是经过运动可以把其中一个图形重合到另一个上面去.在欧氏几何中认为这两个图形是相同的,没有差别的.拓扑学中的同胚映射是更广的一类映射.在同胚的观点下,两个相互同胚的图形认为是相同的,没有差别的.图形在同胚映射下不变的性质叫图形的同胚性质或拓扑不变量.拓扑学就是研究图形的拓扑性质的.

例如，在一个曲面上挖几个洞.在同胚变换下，洞的个数不会变.这就是一个拓扑性质.再如，在一个网络中，顶点个数、弧数以及由弧围成的面数满足一个固定的关系，在拓扑变换下也是不变的.

拓扑学是数学的比较年青而又极为重要的一个分支.著名法国数学家A.魏伊曾经说过，为争取每一个数学家的心灵，拓扑天使和抽象代数恶魔都要角斗.这就是说，第一，拓扑学是优雅而美丽的；第二，整个现代数学是拓扑学和代数学的巧妙的编织物.近年来，拓扑学已深入到物理学、化学、生物学以及心理学中了.

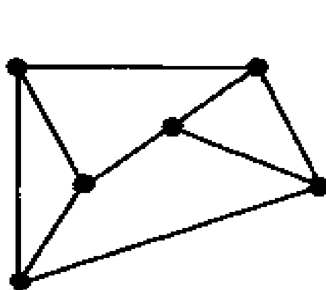
下面再举两个来自拓扑学的例子.

9.1.11　欧拉公式

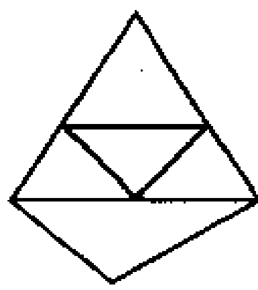
关于网络还有一个特别有用的公式，是欧拉发现的，称为欧拉公式.在拓扑学中它提供了一个基本的不变量.这个公式还可以用来去证明五色定理.它表达了关于网络的三个数之间的一个永恒的关系式.

设 V 表示网络的顶点数，E 表示网络的弧数，这里通常把它称为边数，F 表示面数，也就是由边围成的区域的个数，如图9－24所示.欧拉得到了公式：

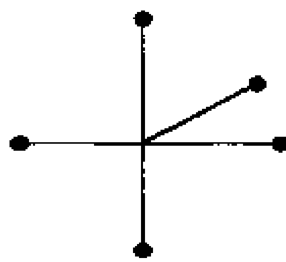
欧拉公式 $V - E + F = 1$.



a) $V=6$ $E=9$ $F=4$



b) $V=7$ $E=11$ $F=5$



c) $V=6$ $E=5$ $F=0$

图 9－24

读者可以用一大堆网络去检验这个公式，你会发现它总是对的.欧拉当时关心的主要是多面体，而不是平面上的网络.图 9 - 25 中给出了多面体的例子.这就解释了为什么使用“顶点”“边”和“面”这样一些名词.我们熟知的五种正多面体是正四面体，立方体，正八面体，正十二面体和正二十面体.它们的顶点、棱、面的数目列表如下：

多面体名称	顶点数	棱 数	面 数
正四面体	4	6	4
立方体	8	12	6
正八面体	6	12	8
正十二面体	20	30	12
正二十面体	12	30	20

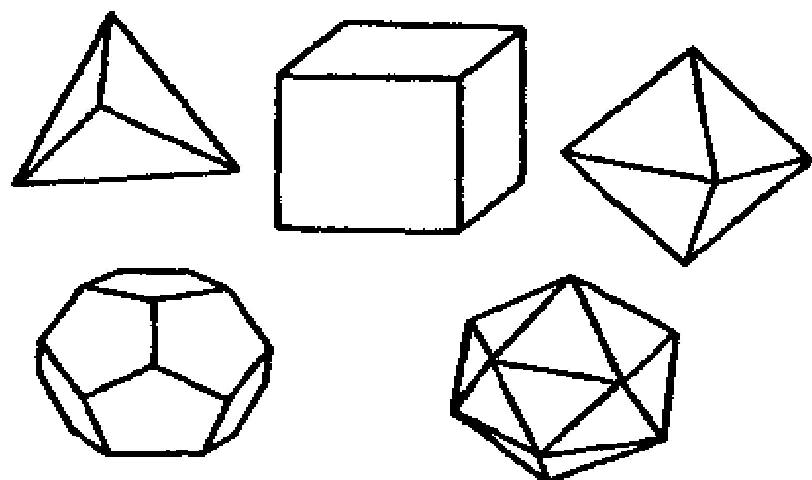


图 9 - 25

对任何一个多面体，欧拉公式取下述形式：

$$V - E + F = 2.$$

这个公式与上面所公式实质上是相同的.事实上，你从多面体上

挖掉一个面,然后将剩下的图形摊开在一张平面上,那么多面体的原来的边就将形成一个连接原来的顶点的网络.反过来,如果有一个网络,你可将它"撑"成缺掉一面的多面体.正是这个缺少了的面解释了网络公式与多面体公式的差别.欧拉公式的证明提供了一种在图论和四色问题的研究中都很有用的研究方法.

欧拉公式的证明 我们从某一网络出发,如图 9-26 所示,去证明公式 $V-E+F=1$.从该网络中去掉一个外边 AB(假如有这样一条外边).这时 E 减少了 1,F 也减少 1,而 V 则保持不变.因此,经过这样的步骤后,$V-E+F$ 保持不变.同样,如果网络中有一个"尾"顶点 B,将这个点连同通向它的边 CB 同时去掉,则 V 减少 1,E 减少 1,而 F 保持不变.在这种情形下,$V-E+F$ 也将保持不变.现在假定你从一个已知网络出发,继续不断地去掉一切可能挪去的外边和尾点,最后你将得到一张只有一个顶点的网络.这时 $V=1$,$E=0$,$F=0$,仍然满足欧拉公式.证毕.

这个证明可以倒过来实行,即从一点开始,去构成任一网络.

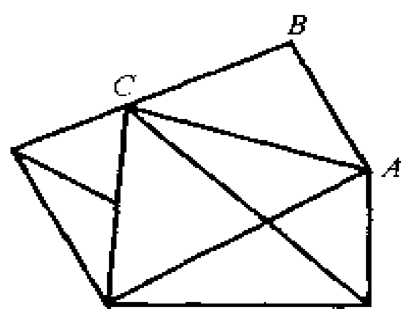


图 9-26

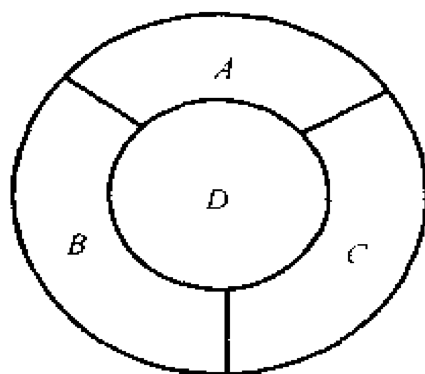


图 9-27

9.1.12 四色问题

1852 年 10 月的一天,英国的一位青年数学家 F. 古色利(Francis

Guthrie)在为一张英国地图着色时发现,似乎只要四种颜色就够了.这当然要满足一个很自然的要求,即任意两个具有公共边界的区域着色不同.但是他证明不了这一事实.于是写信告诉他的弟弟弗雷德里克(Frederick).弗雷德里克转而请教他的数学老师,杰出的英国数学家德·摩根(Augustus de Morgan).

德.摩根很容易地证明了三种颜色是不够的,至少要四种颜色.图 9-27 就说明三种颜色是不够的.德·摩根未能解决这个问题,又把这个问题转给了其他数学家,其中包括著名数学家哈密顿(William Hamilton).但总的说来,这个问题在当时没有引起多大兴趣.1878 年 6 月 13 日,英国数学家凯莱在伦敦数学会上正式提出这个问题,才引起了更大的注意.

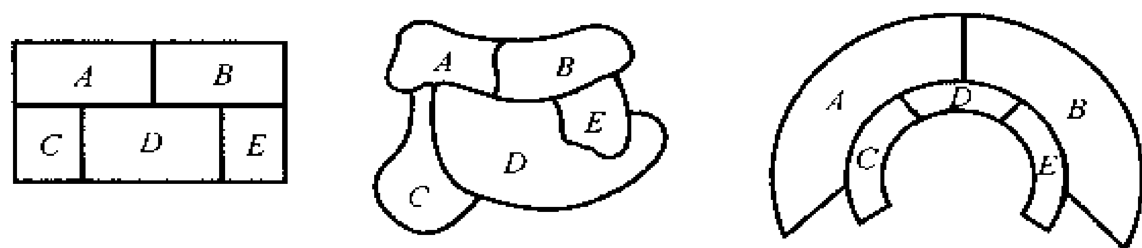


图 9-28

对于地图着色来说,各个地区的实际形状大小并不重要,重要的仅仅是它们的相互位置.例如图 9-28 中的所有地图对地图着色来说都是等价的.从数学上看,问题的实质在于地图的拓扑结构.

一百多年来许多数学家对四色问题进行了大量的研究,获得了一系列成果.1920 年弗兰克林证明了,对于不超过 25 个国家的地图,四色猜想是正确的.1926 年雷诺兹将国家的数目提高到 27 个.1936 年弗兰克林将国家的数目提高到 31 个.1968 年挪威数学家奥雷证明了,不超过 40 个国家的地图可以用四种颜色着色.美国依利诺斯大学的海肯和阿佩尔 1972 年开始人机对话.1976 年 6 月,他们

完成了机器证明.他们使用了3台IBM360型超高速电子计算机,耗时1200小时,终于证明了四色猜想.这是一个非凡的惊人之举,人们盛赞这是计算机革命.当这项成果在1977年发表时,当地邮局特地盖了纪念邮戳"四色足够"(FOUR COLORS SUFFICE).

9.1.13 争论与困惑

四色定理虽然被证明了,但至今仍然有很大争论.多数数学家对这个长长的证明十分不满意,甚至是沮丧的.分析起来争论主要有三个方面.

其一是,证明的可靠性.数学家杜勃在四色定理的证明发表不久就对海肯说:"你的证明在5个月内一定会发现有错误".事实果然如此,错误的确出现了.自论文发表后,阿佩尔等人一直在改正错误.1986年阿佩尔和海肯在"数学信使"杂志上写了一篇解释性文章,说明这些错误都是可以改正的,不是根本性错误.其实人的证明更容易出错误;相对说来,还是机器更少犯错误.国际数学教育委员会在它的第一号研究丛书中指出:"我们不能认为计算机将增加错误证明的数量,恰恰应是反过来."一些数学家认为,四色定理的机器证明向人们展示了计算机的强大威力,它将成为数学家的得力助手,去处理那些极其复杂的定理.

其二是,关于对数学证明的理解.人们对机器证明的主要批评在于这种证明缺乏洞察力,失去了数学的美感.中国著名数学家苏步青说,机器证明"即使是真的,我们总觉得没有什么数学味."人们还是希望在在机器证明的启发下寻找更漂亮,更简短,更富有说服力的证明.

其三是,哲学方面.在什么程度上我们可以说一个依赖于大量的,非人力所能控制的计算机的证明真算一个证明呢?哲学家,斯泰芬·泰缪子寇写道:"如果我们接收四色定理作为一个定理,那么我们就承认改变了'定理'的意义,说得更本质些,改变在'证明'概念下的含义."

数学家真正喜欢的是一个阐明性的证明.这是口味问题而不是逻辑或哲学问题.戴维斯和希尔士评论道,当证明宣布时,许多数学家一定是心中一亮:

"我的第一个反应是,'好极了!他们是如何证明的呢?'我期待着某些光辉的新洞见,一个证明的核心思想中所包含的美会改变我的生活.但是,当我接到下面的答案时,我感到沮丧:'他们把它分成几千种情况,然后在计算机上逐个检查.'我的理由是,'他们只是为了去证明,这毕竟不是一个好问题'."

戴维斯和希尔士还评论道:"对哲学家来说,依赖于机器的可靠性的证明和只依赖人的推理的证明是完全不同的.对数学家来说,推理中的错误是司空见惯的事,他们欢迎计算机,计算机是比他们自己更可靠的计算者."

或许我们可以把最后的话留给海肯,他在接见记者时说:"任何人,任何地方都可沿着这条路线做,可以补充上细节,并检查它们.一个计算机在几小时内可做的细节运算比一个人在一生中希望做的还要多,这一事实并没有改变数学证明的基本概念.改变的不是数学理论,而是数学实践."

习 题

1.可以为图示的网络找到一条最短邮递路线吗?

2.能在图示的房间中找到一条恰恰通过每扇门一次的路线吗?

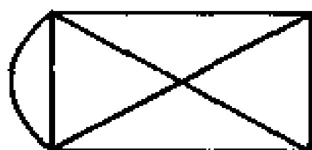
3 能给示的网络找到一条最短邮递路线吗?

4.证明:

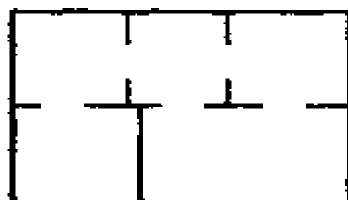
(1)在每次乒乓球比赛中,打过奇数盘的选手共有偶数个(0 算偶数);

(2)参加某次比赛的选手共有 225 人,每人至少打过三盘.证明一定有人打过不止 3 盘.

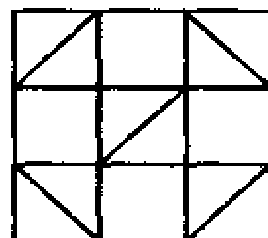
5. 证明对于任何一个网络和它的一个顶点，必能从这点出发走遍整个网络，最后回到原处，并且每条弧走过刚好两遍.



第 1 题



第 2 题



第 3 题

第十章 代数方程式

代数是搞清楚世界上数量关系的智力工具.

A. N. Whitehead

最有价值的科学书籍是作者在书中明白地指出了他所不明白的东西的那些书.遗憾地,这还很少为人们所认识;作者由于掩盖难点,大多害了他的读者.

伽罗瓦

初等数学的主体是代数与几何.其中代数方程主要是围绕一元二次方程展开的.就代数方程式而言,这当然是很不够的,对代数方程的基本方法和结果还缺少一个较为全面的理解.这一讲的目的就在于弥补这方面的不足.我们将介绍以下四个方面的内容.

1)三,四次方程.除了介绍三,四次方程的解法外,还指出,可用根式求解的代数方程只有四次以下的方程.一般的五次以上的代数方程不可用根式求解.

2)代数基本定理.研究代数方程的根的存在问题,指出 n 次代数方程有 n 个根,这是一个非常基本的问题.

3)根的分布问题.根据代数方程的系数来判断它在某个范围的根的个数问题.例如,它有多少正根,多少负根等问题.

4)实根的近似计算.这部分有很强的实用价值.如果我们知道代数方程有实根的话,不用根式也可以求出它的近似解.

有了这些知识,对代数方程的理解就比较全面了.

§10.1 三次方程与四次方程

10.1.1 什么是代数

读者从中学已经熟悉了代数的特征.代数是对字母、字母的表达式进行运算或变换的学问.在初等数学中字母代表数,在近代数学中字母可以代表更广泛的对象,如向量、张量、矩阵、变换等.算术区别于代数的主要依据是,算术仅对具体的数进行运算,而代数对字母进行运算.

但是,什么是代数以及代数的基本问题,随着历史的发展而有改变.代数的发展大致分为三个时期.第一个时期从9世纪的花拉子米始,到16世纪止.这个时期人们把代数看成为对字母进行运算,关于字母公式的变换以及关于代数方程式的学问.这些就是目前中学代数的内容.第二个时期从16世纪开始到19世纪,这时意大利数学家解出了三次方程和四次方程.由此人们开始研究解更高次的代数方程.代数的中心问题逐渐变为代数方程式的理论了.19世纪谢尔的两卷本的代数问世,在这部书中代数被定义为方程式论.这在当时是个创举.在这部书中第一次讲述了代数方程式论的顶峰——伽罗瓦理论.在第二个时期内,行列式与矩阵的理论,二次型与变换的理论,特别是不变量的理论等代数工具也发展起来了.在这个时期内群论及不变量的理论的发展对几何学的发展起了重大影响.第三个时期从上世纪末到本世纪.这时在力学、物理以及数学本身越来越频繁地研究到一些对象,对这些对象也要考虑加法、减法,有时要考虑乘法和除法.但这些运算满足的运算规律不同于有理数.这些对象中有矩阵、张量、旋量、超复数等.这样人们就不得不考虑某种更一般的集合,在这种集合中有某种运算,并满足一定的运算法则.这就是说,我们不得不考虑某种代数系统.这样一来,代数的目的是研究各种代数系统.这就是公理化的,或抽象的代数.说它是抽象的,是因为所考虑

的代数系统是用字母表示的.说它是公理化的,是因为它只遵从作为它的基础的那些公理.有趣的是这样的代数系统无论就数学本身而言,或就它的应用而言都具有巨大的意义.20 世纪 30 年代,范·德·瓦尔登的名著《代数学》对阐述什么是代数的第三个观点起了巨大的作用.

最近几十年来计算机的使用正在改变着代数的面貌,为代数学提供了许多新的特殊的课题.

10.1.2　二次方程

先讲一点历史,看古人是如何解二次方程的.20 世纪的考古工作发现,早在公元前 1700 年,居住在美索不达米亚的人们已经有了高度发展的数学文化,其中包括 60 进位制和勾股定理的知识.他们知道勾股定理是在毕达哥拉斯以前一千年.他们已经有了解二次方程的成法.巴比伦人将二次方程的解法化为一种正规形式,其正规形式是,"已知两数的和与积求此两数".用现代的代数语言来叙述就是,给定两个数 p 和 q,并已知 $xy = p$, $x + y = q$,求 x, y.巴比伦人用下述五个步骤求这两个数:

1.取 p 的一半;

2.将此数平方;

3.从中再减去 q;

4.对所得结果的开平方;

5.再加 p 的一半得出所求两数中的一数;从 p 中减去这个数得出另一数.

例　设 $xy = q = 21$, $x + y = p = 10$,求 x, y.

解　按上面的步骤有:

1)5; 2)25; 3)4; 4)2; 5)$x = 7$, $y = 3$(或 $x = 3$, $y = 7$).

巴比伦人的正规形式,用现代语言来说就是一元二次方程.事实上,

$$px = (x+y)x = x^2 + xy = x^2 + q \Leftrightarrow x^2 - px + q = 0.$$

x 是二次方程的解.根据对称性,y 也是二次方程的解.

但是巴比伦人还不能把所有的二次方程都化为正规形式,因为在那个时代还没有负数的概念.负数的概念只是在几个世纪以前才诞生.

巴比伦人的正规形式的五个步骤用近代的代数语言来写,就是

$$x = \sqrt{\left(\frac{p}{2}\right)^2 - q} + \frac{p}{2}, y = p - x.$$

这可以化为我们更熟悉的形式

$$x, y = \frac{p \pm \sqrt{p^2 - 4q}}{2}.$$

他们是如何推出这一公式的呢?我们现在无从知道,因为那样遥远的年代的遗存物是太少了.

10.1.3 韦达公式

知道了二次方程的两个根 x_1, x_2 就可将它分解因式.我们有

$$x^2 + px + q = (x - x_1)(x - x_2) = 0.$$

由此不难得出著名的韦达公式:

$$x_1 + x_2 = -p, x_1 \cdot x_2 = q.$$

利用代数基本定理我们可以得到更一般的公式.

代数基本定理 设

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n \tag{1}$$

是一个给定的 n 次多项式,它的系数 $a_1, a_2, \cdots, a_n$ 是实数或复数,那么方程

$$f(x) = 0$$

至少有一个实数或复数根.

有了代数基本定理,我们就可以把 n 次多项式 $f(x)$ 分解成一次因式的连乘积,即

$$f(x)=(x-x_1)(x-x_2)\cdots(x-x_n),\tag{2}$$

这里 $x_1,x_2,\cdots,x_n$ 实数或复数.

事实上,设 x_1 是方程的一个根,用 $(x-x_1)$ 去除 $f(x)$,由于除式是一次的,所以余数就是一个常数 R,即我们有恒等式

$$f(x)=(x-x_1)f_1(x)+R.$$

式中 $f_1(x)$ 是一个 $n-1$ 次多项式,而 R 是常数.把 x_1 代入上式,就得到

$$f(x_1)=(x_1-x_1)f_1(x_1)+R=R=0.$$

因为 x_1 是 $f(x)$ 的一个根.这就是说,$(x-x_1)$ 能整除此多项式,所以

$$f(x)=(x-x_1)f_1(x).$$

同样的道理,我们有

$$f_1(x)=(x-x_2)f_2(x).$$

n 次分解之后,我们得到(2)式.把(2)式乘开,并比较系数就得到 n 次方程的韦达公式:

$$\begin{aligned}
a_1&=-(x_1+x_2+\cdots+x_n),\\
a_2&=x_1x_2+x_1x_3+\cdots+x_1x_n+x_2x_3+\cdots+x_{n-1}x_n,\\
&\cdots\cdots\cdots\cdots\cdots,\\
a_n&=(-1)^nx_1x_2\cdots x_n.
\end{aligned}$$

当 $n=2$ 时,就是我们熟知的二次方程的根与系数的关系.对三次方程

$$x^3+a_1x^2+a_2x+a_3=0.$$

我们有

$$\begin{aligned}
a_1&=-(x_1+x_2+x_3),\\
a_2&=x_1x_2+x_2x_3+x_3x_1,\\
a_3&=-x_1x_2x_3.
\end{aligned}\tag{4}$$

这就是三次方程的韦达公式,下面要用.

10.1.4　三次方程

现在我们讨论三次方程的解法. 一般中学数学中不包含这部分内容. 用配方法解二次方程早在古代巴比伦就已经知道了. 高于二次的方程就是另一回事了. 解一般的三次方程要困难得多. 这使得许多古代数学家的努力都归于失败. 直到 16 世纪初的意大利的文艺复兴时代，这个问题才被意大利数学家所解决.

解三次方程的步骤分为三步：1）将一般方程化为缺项的三次方程；2）解缺项的三次方程；3）解的确定.

设一元三次方程为

$$y^3 + a_1 y^2 + a_2 y + a_3 = 0. \tag{5}$$

1）我们首先证明，它可以化为缺项的三次方程

$$x^3 + px + q = 0. \tag{6}$$

事实上，作变换 $y = x - \dfrac{a_1}{3}$，把它代入(5)，得

$$\left(x - \frac{a_1}{3}\right)^3 + a_1\left(x - \frac{a_1}{3}\right)^3 + a_2\left(x - \frac{a_1}{3}\right)^3 + a_3$$

$$= x^3 - 3 \cdot x^2 \cdot \frac{a_1}{3} + \cdots + a_1 x^2 + \cdots.$$

式中"…"表示 x 的一次项和零次项各项. 可见在上式中含 x^2 的项相互抵销了. 再合并同类项，就得到(6). 其中

$$p = -\frac{a_1^2}{3} + a_2, \quad q = \frac{2a_1^3}{27} - \frac{a_1 a_2}{3} + a_3.$$

2）现在我们来解方程(6). 解法虽然不长，但是相当巧妙. 方法是引进两个未知量 u 和 v 代表一个未知量 x. 具体是这样作的：设 $x = u + v$. u，v 代表两个新的未知量. 把它代入方程(6)，得到

$$(u + v)^3 + p(u + v) + q = 0.$$

展开第一项，得到

$$u^3 + 3u^2 v + 3uv^2 + v^3 + p(u + v) + q = 0.$$

合并同类项，得到

$$u^3 + v^3 + q + 3uv(u + v) + p(u + v) = 0.$$

进而得

$$(u^3 + v^3 + q) + (3uv + p)(u + v) = 0. \tag{7}$$

因为我们用了两个未知量 u 和 v 代替一个未知量 x，所以还可以再加一个条件. 今要求

$$3uv + p = 0 \Leftrightarrow uv = -\frac{p}{3}. \tag{8}$$

这样一来，(7) 式变成了两个方程：

$$u^3 + v^3 = -q, \quad u^3 v^3 = -\frac{p^3}{27}.$$

从这两个方程不难看出，u^3 和 v^3 是二次方程

$$z^2 + qz - \frac{p^3}{27} = 0$$

的两个根，解这个二次方程得

$$z = -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

由此，得出

$$u = \sqrt[3]{z_1} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},$$

$$v = \sqrt[3]{z_2} = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{9}$$

这样，我们就完成了缺项的三次方程的解法：

$$x = u + v = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{10}$$

3) 解的确定. 因为一个立方根在复数域中有三个值，所以(9) 式给予 u 三个值和 v 三个值. 互相搭配起来共有 9 个值，而三次方程只

有三个根,因此利用(10)式求 x 的值时,不能取 u,v 的值的任意组合,必须使它们满足(8)式才是解.

设 u_1 是 u 的三个值中的任一个.如本书上册第二讲复数一节所指出的,u 的另外两个值可用 1 的立方根 ω 与 ω^2 乘 u_1 来得到

$$u_2 = u_1\omega,\ u_3 = u_1\omega^2,$$

$$\omega = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2},\ \omega^2 = -\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}.$$

用 v_1 表示 v 的三个值中乘 u_1 满足(8)式的那个值,即 $u_1v_1 = -\dfrac{p}{3}$. v 的另外两个值是:$v_2 = v_1\omega, v_3 = v_1\omega^2$.我们来看看,分别与 u_2, u_3 相对应的 v 是那两个.因为 $\omega^3 = 1$,

$$u_2v_3 = u_1\omega \cdot v_1\omega^2 = u_1v_1\omega^3 = u_1v_1 = -\frac{p}{3},$$

$$u_3v_2 = u_1\omega^2 \cdot v_1\omega = u_1v_1\omega^3 = u_1v_1 = -\frac{p}{3}.$$

所以 u_2 与 v_3 对应,u_3 与 v_2 对应.这样一来,方程(6)的三个根是

$$x_1 = u_1 + v_1,$$

$$\begin{aligned} x_2 &= u_2 + v_3 = u_1\omega + v_1\omega^2 \\ &= -\frac{1}{2}(u_1 + v_1) + \mathrm{i}\frac{\sqrt{3}}{3}(u_1 - v_1), \end{aligned} \tag{11}$$

$$\begin{aligned} x_3 &= u_3 + v_2 = u_1\omega^2 + v_1\omega \\ &= -\frac{1}{2}(u_1 + v_1) - \mathrm{i}\frac{\sqrt{3}}{3}(u_1 - v_1). \end{aligned}$$

至此,我们完成了对三次方程的求解.公式(11)称为卡尔达诺公式.

例　解三次方程 $x^3 - 6x + 6 = 0$.

解　我们用卡尔达诺公式求解.在此有 $p = -6, q = 6$.代入(9),得

$$u = \sqrt[3]{-3 + \sqrt{9-8}} = \sqrt[3]{-2}.$$

为方便计，把 u_1 取为 -2 的立方根的实数值，即令 $u_1=-\sqrt[3]{2}$. 和这个 u_1 对应的 v_1 是

$$v_1=-\frac{p}{3u_1}=-\frac{2}{\sqrt[3]{2}}=-\sqrt[3]{4}.$$

利用公式(11)，我们可以得到方程的三个根：

$$x_1=-\sqrt[3]{2}-\sqrt[3]{4},$$

$$x_2=\frac{1}{2}(\sqrt[3]{2}+\sqrt[3]{4})+\mathrm{i}\frac{\sqrt{3}}{2}(\sqrt[3]{4}-\sqrt[3]{2}),$$

$$x_3=\frac{1}{2}(\sqrt[3]{2}+\sqrt[3]{4})-\mathrm{i}\frac{\sqrt{3}}{2}(\sqrt[3]{4}-\sqrt[3]{2}).$$

10.1.5　实系数的三次方程

我们日常遇到的三次方程大多是实系数的，现在我们对实系数的三次方程作一些详细的讨论. 二次方程的根的性质是通过它的判别式来讨论的. 我们来引进三次方程的判别式. 令

$$D=\frac{q^2}{4}+\frac{p^3}{27}.$$

我们称它为三次方程(6)的判别式. 它决定了方程的根的性质. 分三种情况进行讨论：1) $D>0$；2) $D=0$；3) $D<0$. 这时

$$u=\sqrt[3]{-\frac{q}{2}+\sqrt{D}},\quad v=\sqrt[3]{-\frac{q}{2}-\sqrt{D}}.$$

1) $D>0$ 的情况. 这时

$$-\frac{q}{2}+\sqrt{D},\qquad -\frac{q}{2}-\sqrt{D}$$

都是实数且不相等. 在这种情况下可令 u_1 等于 $-\frac{q}{2}+\sqrt{D}$ 的实数值的立方根. 因为 u_1v_1 必须等于 $-\frac{p}{3}$ 实数，所以 v_1 也必须等于 $-\frac{q}{2}-\sqrt{D}$ 的实数值立方根. 由此，$x_1=u_1+v_1$ 是实数. 剩下的两个根是复数：

$$x_2 = -\frac{1}{2}(u_1 + v_1) + \mathrm{i}\frac{\sqrt{3}}{2}(u_1 - v_1),$$

$$x_3 = -\frac{1}{2}(u_1 + v_1) - \mathrm{i}\frac{\sqrt{3}}{2}(u_1 - v_1).$$

这是一对共轭复数.这就证明了,当 $D > 0$ 是,方程(6) 有一个实根和两个互为共轭的复根.

2) $D = 0$ 的情形.这时 $u^3 = v^2 = -\frac{q}{2}$,所以

$$x_1 = -2\sqrt[3]{\sqrt{\frac{q}{2}}}, \quad x_2 = x_3 = \sqrt[3]{\sqrt{\frac{q}{2}}}.$$

因此,方程(6) 的三个根都是实根,而且有两个根相等.

3) $D < 0$ 的情形.这时立方根内的数不再是实数而是复数.从而 u,v 也是复数.我们指出,u 和 v 一定是共轭的.根据开 n 次方根的规则:$\left|\sqrt[n]{z}\right| = \sqrt[n]{|z|}$,我们有

$$|u| = \left|\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}\right| = \sqrt[3]{\left|-\frac{q}{2} + \mathrm{i}\sqrt{\frac{q^2}{4} + \frac{p^3}{27}}\right|}$$

$$= \sqrt[3]{\sqrt{\frac{q^2}{2} - \frac{q^2}{4} - \frac{p^3}{27}}} = \sqrt[3]{\sqrt{-\frac{p^3}{27}}} = = \sqrt{-\frac{p}{3}}$$

现在容易证明 u 和 v 共轭了:

$$v = -\frac{p}{3u} = -\frac{p\bar{u}}{3u\bar{u}} = -\frac{p\bar{u}}{3|u|^2} = -\frac{p\bar{u}}{3\left(-\frac{p}{3}\right)} = \bar{u}.$$

现在我们来给出方程(6) 的三个根.设 $u_1 = a + \mathrm{i}b$ 是 u 的任意一个值,从而 $v_1 = a - \mathrm{i}b$.因此,

$$x_1 = u_1 + v_1 = 2a,$$

$$x_2 = -\frac{1}{2}(u_1 + v_1) + \mathrm{i}\frac{\sqrt{3}}{2}(u_1 - v_1) = -a - b\sqrt{3},$$

$$x_3 = -\frac{1}{2}(u_1 + v_1) - \mathrm{i}\frac{\sqrt{3}}{2}(u_1 - v_1) = -a + b\sqrt{3}.$$

换句话说，所得的三个根都是实根.这三个实根彼此互异.

在16世纪，负数开方被认为是不可能的，因为当时还没有复数的概念.所以$D<0$的情形使当时的数学家感到困惑.他们不知道为什么会从不可能的运算中得到实根.他们花了很大的力气企图消去卡尔达诺公式中的虚数性，但都归于失败.顺便指出，$D<0$的情形还与三等分任意角相关.

10.1.6 卡尔达诺公式小史

卡尔达诺公式最先刊登于1545年卡尔达诺出版的著作《大法》(Ars magna)里.卡尔达诺(Cardano Jerome，1501—1576)是意大利米兰的数学和物理教授.他的这种方法得之于意大利数学家塔尔塔利亚(Tartaglia Niccolo，1500—1557).这里有一段有趣的故事.原来这一问题最初是由意大利数学家齐波·费罗(Scipione dal Ferro 1465～1526)解决的.但他没有发表他的解法.按照当时的风气，人们常把所得到的发现保密，而向对手提出挑战，要他们解决同样的问题.这种做法在"不发表就发霉"的今天是不可思议的.费罗对他的方法终生保密，直到弥留之际才将他的方法传给了他的学生安东尼奥·菲奥尔(Antonio Maria Fior).费罗去世后，菲奥尔向当时意大利最大的数学家之一的塔尔塔利亚提出挑战，要他解出30个三次方程.塔尔塔利亚起而应战，并且用8天时间结束了这场竞赛，解出了对手提出的所有30个方程，得到了解缺项三次方程的一般方法.

当卡尔达诺获悉这一发明后，就央求塔尔塔利亚将密诀告诉他，并发誓对此保守秘密.在卡尔达诺的恳求下，塔尔塔利亚把他的方法写成一首晦涩的诗告诉了卡尔达诺.但是卡尔达诺背弃了他的诺言，而将方法发表了.

10.1.7 三次方程解法总结

三次方程的成功解出为四次方程的解出开辟了成功之路.所以值得将三次方程的解法再作一小结：

1）将完全三次方程化为缺项三次方程；

2）引进一对辅助变量 u,v 及一个辅助的二次方程；

3）解二次方程得到 u^3、v^3，由此得到缺项三次方程的解；

4）解的确定．

值得注意的是，要解一个三次方程，必须先解一个二次方程．这个方法启发了意大利数学家费拉里（Lodovico Ferrali 1522 ~ 1565），他很快就给出了一般四次方程的解法，并发表在卡尔达诺的“大法”中．

10.1.8　四次方程

四次方程的解法比起三次方程要复杂多了，但基本思想类似于三次方程．其主要步骤如下．

1）将完全的四次方程化为缺项的四次方程．

2）引进三个辅助变量 u,v,w，得到 u,v,w 的三组关系式，并引出一个辅助的三次方程．

3）解三次方程得到 u,v,w．

4）解的确定．u,v,w 的值不能任意搭配，需满足有关关系式．由此得到缺项四次方程的解．

总之，将缺项的四次方程化为三次方程、解出三次方程后，再求出四次方各程的根．整个过程是：

解四次方程化为解三次方程，解三次方程化为第二次方程．

现在我们来研究四次方程的解法．设一元四次方程为

$$x^4+a_1x^3+a_2x^2+a_3x+a_4=0. \tag{13}$$

首先把它化为不含 x^3 项的三次方程：

$$x^3+qx^2+rx+s=0. \tag{14}$$

事实上，令 $y=x-\frac{a_1}{4}$，代入(13)式得

$$\left(x-\frac{a_1}{4}\right)^4+a_1\left(x-\frac{a_1}{4}\right)^3+\cdots=0,$$

$$x^4 - 4x^3 \cdot \frac{a_1}{4} + a_1 x^3 + \cdots = 0,$$

式中“…”表示 x 的幂低于的项. 由此可见,含 x^3 的项互相抵消. 这样一来,问题化为解缺项的四次方程(14)了.

解缺项的四次方程比缺项的三次方程要复杂,需要引进的辅助变量不是两个而是三个. 设它们是 u, v, w. 另外,引进的辅助方程也不再是二次方程,而是三次方程了. 下面我们就来求解方程(14).

令 $x = u + v + w$,于是有

$$x^2 = u^2 + v^2 + w^2 + 2(uv + vw + wu),$$

$$\begin{aligned} x^4 = &(u^2 + v^2 + w^2)^2 \\ &+ 4(u^2 + v^2 + w^2)(uv + vw + wu) \\ &+ 4(uv + vw + wu)^2. \end{aligned}$$

把所有这些结果代入(14),得

$$\begin{aligned} &(u^2 + v^2 + w^2)^2 + 4(u^2 + v^2 + w^2)(uv + vw + wu) \\ &+ 4(uv + vw + wu)^2 + q(u^2 + v^2 + w^2) \\ &+ 2q(uv + vw + wu) + r(u + v + w) + s = 0. \end{aligned}$$

注意到

$$\begin{aligned} &4(uv + vw + wu)^2 \\ &= 4[(u^2v^2 + v^2w^2 + w^2u^2) + 2uvw(u + v + w)], \end{aligned}$$

从而得到

$$\begin{aligned} &(u^2 + v^2 + w^2)^2 + 2(uv + vw + wu)[2(u^2 + v^2 + w^2) + q] \\ &+ q(u^2 + v^2 + w^2) + (8uvw + r)(u + v + w) \\ &+ 4(u^2v^2 + v^2w^2 + w^2u^2) + s = 0. \end{aligned} \tag{15}$$

u, v, w 是三个变数,而这里只有一个方程,要通过 u, v, w 来确定 x 还必须增加两个条件. 为此设

$$2(u^2 + v^2 + w^2) + q = 0 \Leftrightarrow u^2 + v^2 + w^2 = -\frac{q}{2} \tag{16}$$

和

$$8uvw + r = 0 \Leftrightarrow uvw = -\frac{r}{8}. \tag{17}$$

这时(15)变为

$$(u^2+v^2+w^2)^2 + q(u^2+v^2+w^2) + 4(u^2v^2+v^2w^2+w^2u^2) + s = 0.$$

再把(16)代入,可得

$$\frac{q^2}{4} - \frac{q^2}{2} + 4(u^2v^2+v^2w^2+w^2u^2) + s = 0.$$

从而

$$u^2v^2 + v^2w^2 + w^2u^2 = \frac{q^2-4s}{16}. \tag{18}$$

由(17)得

$$u^2v^2w^2 = \frac{r^2}{64}. \tag{19}$$

将(16),(18),(19)结合起来,利用三次方程的韦达公式可知,u^2,v^2,w^2 是下面的三次方程的根:

$$z^3 + \frac{q}{2}z^2 + \frac{q^2-4s}{16}z - \frac{r^2}{64} = 0. \tag{20}$$

若这个三次方程的根是 z_1, z_2, z_3,则

$$u = \pm\sqrt{z_1}, \quad v = \pm\sqrt{z_2}, \quad w = \pm\sqrt{z_3}.$$

这时 $x = u + v + w$ 有 8 种可能的结合,但由于有条件(17)的限制,所以实际上只有 4 种结合,这就是四次方程的四个根.

我们看到了,解四次方程要预先解一个形如(20)的三次方程.

一个重要的事实是,一次方程,二次方程,三次方程和四次方程的根都是通过系数的四则运算,乘方与开方等运算来表示的.这就诱使人们借助同样的运算去寻求五次以上方程的求根公式.但是这一努力失败了.

还有一点值得指出,我国对高次方程的研究也开始得很早.早在唐朝王孝通著的《辑古算经》就记载有缺项的三次方程,并说明“以

立方除之”.到了13世纪的金元时期,李冶(1192—1279),秦九韶(1247),杨辉(1261—1275),朱世杰(1303)等,都曾对高次方程提出解法.当时用算筹可以解出十次方程.但在我国古代数学偏重于应用,只讨论正根不讨论负根,自然更没有复根.不过在13世纪我们已经能求出高到十次方程的正根,这仍然值得我们自豪.

10.1.9 五次以上的代数方程

五次以上的代数方程如何解?是不是和二次,三次,四次方程一样?它们的根是否可以用系数的四则运算,乘方与开方等运算表示出来,只不过是技巧更高超,表达式更复杂呢?产生这种想法是很自然的.代数发展史本身就是一个很好的说明.

在意大利数学家成功地解出了三次方程和四次方程后,极大地鼓舞了当时的数学家,他们立刻开始研究高次方程的解法.试图用根式解出五次、六次乃至更高次的方程.这种努力持续了两个半世纪之久,而没有获得成功.自然界的一个普遍法则是量变引起质变.方程的次数高到一定程度(这里是五),原来的方法就失效了.当时的数学家做梦也没有想到他们的努力是徒劳的.直到18世纪的后半叶人们才意识到这一结局.法国数学家拉格朗日在1770年—1771年发表的长文“关于代数方程解法的思考”中指出,用代数运算解一般的高次方程($n>4$)看来是不可能的.这里的代数运算指的是加、减、乘、除、乘方(指数是整数)与开方这六种运算.他说,或者这个问题超出了人类的智力范围,或者是根的表达方式一定不同于当时所知道的一切.后面这一猜测道出了问题的关键所在.

拉格朗日的方法尽管很少成功,但他确实给出了洞察$n<5$时成功,$n>4$时失败的道理.这种洞察力为阿贝尔和伽罗瓦所利用.

受拉格朗日的影响,鲁菲尼(Paolo Ruffini,1765—1822)在1799到1813年之间作过好几种尝试,要证明四次以上的方程不能用代数方法解出,但他的努力不甚成功.

1824年,当天才的挪威青年数学家阿贝尔(Abel,1802—1829)

的著作出版时引起了所有数学家的惊奇.如果方程的次数 $n \geqslant 5$,并且将方程式的系数看成字母,那么任何一个由这些字母组成的根式都不可能是方程的根.原来一切国家的最伟大的数学家三个世纪以来用根式去解五次以上的方程所以不能成功,是因为这个问题根本没有解.

1858 年埃尔米特(Charles Hermite,1822—1902)用椭圆函数给出了一般五次方程的解.后来他又成功地借助富克斯函数将一般 n 次方程的根用它的系数表示出来.

阿贝尔的工作之后,情况是这样的:虽然高于四次的一般方程不能用根式求解,但仍有很多特殊的方程,如我们曾经讨论过的二项方程 $x^p = a$ 和阿贝尔方程都仍可用根式求解.现在的问题是确定哪些方程可用根式求解.刚刚由阿贝尔开始的这个任务由天才的法国数学家伽罗瓦(Evariste Galois,1811—1832)担当起来了.他找出了方程能用根式解出的充分和必要条件.

伽罗瓦是数学史上罕见的天才之一,他 20 岁就因决斗而身亡.他 15 岁进入巴黎的一所有名的公立学校,并开始研究数学.他仔细研究了拉格朗日,高斯,柯西和阿贝尔的著作.在学校的第一年他发表了四篇文章.1829 年,他把解方程的两篇文章呈送科学院.文章转给了柯西,可惜柯西把它们遗失了.1830 年他交给科学院另一篇仔细写成的关于他的研究的文章.这篇文章送到了傅里叶那里,不久傅里叶就去世了,这篇文章也被遗失了.在泊松提议下,1831 年伽罗瓦就他的研究写了一篇新文章"关于用根式解方程的可解性条件",遗憾的是,泊松看不懂,以难以理解为由,又将稿子退回,并劝告再写详细些.在伽罗瓦决斗的前夜,关于他的研究,他匆忙地写了一份说明,托给了他的朋友 A.车若里尔(August Cheralier).这个说明保存了下来.1846 年,刘维尔在"数学杂志"上发表了伽罗瓦的部分文章,并包括对 1831 年文章的一个修订.1870 年,法国数学家若尔当发表了关于伽罗瓦理论的头一个全面而清楚的介绍.这使得伽罗瓦的发现完

全为人们所理解，并且确立了他在数学史上的地位. 伽罗瓦的名字永存于“伽罗瓦域”、“伽罗瓦群”、“伽罗瓦理论”中，它们都是近世代数所研究的最重要的课题，他的工作构成了 19 世纪数学的最杰出的成就之一.

习　题

解下列方程：

1. $x^3+2x-1=0$，　　2. $x^3+3x^2-4=0$，

3. $x^3-3x^2+5x-1=0$，　4. $x^4+3x^2-5x-3=0$.

§10.2　代数基本定理

10.2.1　引言

上节我们考察了代数方程的求解问题，并给出了三次方程和四次方程的解法. 我们还介绍了数学家企图用根式求解 n 次方程的奋斗史. 经过三个世纪的努力终于发现，当 $n>4$ 时那是不可能的. 问题的确是很困难、很深刻的. 这个问题的研究引出非常重要的新思想，这些思想不仅对代数而且对整个数学都具有划时代的意义.

至于谈到方程的实际解法，需要指出的是，用根号解方程的方法对一切方程来说是远远不够的，除了二次方程以外，即使能用根号解也由于方法的复杂性而具有较少的实用价值. 因此，数学家对代数方程理论的研究早就着手在以下三个方面进行工作：

1）关于根的存在问题；

2）不解出方程而根据系数来判断方程的根的性质；

3）关于方程的根的近似计算问题.

我们首先要证明的是，代数基本定理：每一个实系数或复系数的

n 次代数方程至少有一个实根或复根.

非代数的方程就不一定有根.

例　方程 $a^x = 0$ 既没有实根也没有复根.其中 a 是任一实数或复数.

这个定理是整个数学中最重要的定理之一,它在代数学中起着基石的作用,所以称它为代数基本定理.代数基本定理的证明是困难的,经过几代数学家的努力才得到它的严格证明.不无惊奇的是,就其本质而言,证明方法不是代数的而是分析的.代数基本定理的任何一个证明都用到了相当深刻的分析结果.因而它的任何严格证明只能出现在分析的严密化之后.这之前的任何证明都必然有这样或那样的缺陷.代数基本定理的第一个证明是达朗贝尔给出的.他用到数学分析的一个命题:定义在有限闭区间上的连续函数一定在某一点取得最小值.这个命题的严格证明是在18世纪的后半叶才得到的,即在达朗贝尔的研究一百年后才得到的.

1799年高斯对代数基本定理给出了他的第一个证明.他的证明依赖于对复数的承认,从而对巩固复数的地位作出了贡献.在高斯以及高斯以前那些年代,尽管复数获得了许多卓有成效的应用,人们依然对它怀有疑惧,在当时数学家的眼中复数仍然不是数学大家庭中的合法成员.高斯的证明也不是完全严格的.除此之外,高斯还给出了这个定理的三个别的证明.这里我们给出这个定理的一个较初等的证明.

10.2.2　代数基本定理的证明

考虑复数序列

$$z_1, z_2, \cdots, z_n, \cdots.$$

定义　称一个复数序列 $\{z_n\}$ 是有界的,如果存在一个正数 $M > 0$,使得 $|z_n| < M$ 对一切 n 成立.

定义　称复数序列 $\{z_n\}$ 以 a 为极限,若 $\forall \varepsilon > 0, \exists N > 0$,当

$n > N$ 时,

$$|z_n - a| < \varepsilon.$$

记为

$$\lim_{n\to\infty} z_n = a.$$

易见,

$$|\mathrm{Re}z_n - \mathrm{Re}a| \leqslant |z_n - a|, \quad |\mathrm{Im}z_n - \mathrm{Im}a| \leqslant |z_n - a|$$

和

$$|z_n - a| \leqslant |\mathrm{Re}\, z_n - \mathrm{Re}\, a| + |\mathrm{Im}\, z_n - \mathrm{Im}\, a|.$$

将这两个式子结合起来,立刻可看出,复数序列的收敛问题可化为实数序列的收敛问题。

引理1(波尔查诺-魏尔斯特拉斯) 一个有界的复数序列有收敛的子序列.

我们只需证明,一个有界的实数序列有收敛的子序列就行了.关于实数序列的定理,其严格证明放在微积分部分.

考虑 n 次多项式

$$f(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_{n-1} z + a_n. \tag{1}$$

现在我们来着手证明代数基本定理.证明分为两步:

1) 在复平面上存在一点 z_0,使得对复平面上的任何 z 都有 $|f(z_0)| \leqslant |f(z)|$,即多项式的模在复平面的某一点处取得最小值;

2) 若 z_0 使得 $|f(z_0)|$ 取得最小值,则 $f(z_0) = 0$.

多项式模的最小值定理

我们先证明第一部分.为此需要下面的引理.

引理2 设多项式 $f(z)$ 的次数 $n \geqslant 1$,那么对任意给定的正数 M,都存在一个实数 $R > 0$,使得对一切 $|z| > R$,都有

$$|f(z)| \geqslant M.$$

证 我们用归纳法证明这一引理.对多项式的次数 n 作归纳

法.

首先,设 $n=1$,$f(z)=a+bz$,$b\neq 0$.那么

$$|f(z)|=|a+bz|\geqslant|bz|-|a|=|b||z|-|a|.$$

对于给定的 M,取 $R=(M+|a|)/|b|$,于是当 $|z|>R$ 时,$|f(z)|\geqslant M$.

其次,假定引理对 $k-1$ 次多项式成立.设 $f(z)$ 的次数是 k,则 $f(z)=a+zf_1(z)$,$f_1(z)$ 是 $k-1$ 次多项式.对于给定的 M,取 R(在 $R>1$ 的范围内取值),使得当时 $|z|>R$ 时,

$$|f_1(z)|>M+|a|.$$

于是当 $|z|>R$ 时,

$$|f(z)|=|a+zf_1(z)|\geqslant|z||f_1(z)|-|a|\geqslant|f_1(z)|-|a|,$$
$$M+|a|-|a|=M.$$

证毕.

现在我们可以证明多项式模的最小值定理了.

定理 1 存在复数 z_0,使 $|f(z_0)|$ 最小.

证 设 $|f(0)|=g$,并任取 $G>g$.根据引理 2,存在一个正数 R,当 $|z|>R$ 时,$|f(z)|>G$,这样一来,$|f(z)|$ 不会在 $|z|\leqslant R$ 外取得最小值.我们可将注意力集中于 $|z|\leqslant R$ 内.

如果 $g=0$,则显然 $|f(z)|$ 在点 $z=0$ 处达到最小值,点 0 就是我们所找的 z_0,定理自然得证.因而我们考虑 $g>0$ 的情况.

如果 $g>0$,并且对一切 z 都有 $|f(z)|\geqslant g$,那么 $|f(z)|$ 也在 0 点处有最小值,0 就是我们所找的 z_0,定理自然成立.

在一般情况下,$g>0$,且 g 不是最小值,我们要找一点 z_0,使 $|f(z)|$ 最小.

用什么方法去找这一点呢?用引理 1.这就是说,我的方法是分析的而不是代数的.具体方法如下.

既然 g 不是 $|f(z)|$ 的最小值,$|f(z)|$ 的最小值就一定落在区间 $[0,g]$ 内,把 $[0,g]$ 分成 n 等分:

$$0,\frac{g}{n},\frac{2g}{n},\cdots,\frac{ng}{n}=g\,(n=2,3,4\cdots).$$

这是一个递增的序列.由于对一切 z 都有 $|f(z)|\geqslant 0$,而又存在 z 使得 $|f(z)|<g$,所以在上面的序列中一定有一个最大的 $\frac{ig}{n}$,使得对一切的 z 都有 $|f(z)|\geqslant\frac{ig}{n}$,而存在 z_n 使 $|f(z_n)|<\frac{i+1}{n}g$.记 $C_n=\frac{ig}{n}$,$C'_n=\frac{i+1}{n}g$.

由于 $|z|>R$ 时,$|f(z)|>G>g$,所以对一切的 n 都有 $|z_n|\leqslant R$.这就是说,$\{z_n\}$ 构成有界的点列.根据引理 1,从中可以选出一个收敛的子序列 $\{z_{n_k}\}$,它收敛到点 z_0:

$$\lim_{k\to\infty}z_{n_k}=z_0.$$

我们来证明,在 z_0 处 $|f(z)|$ 有最小值.事实上,设 z 是任意一点,这时,

$$\begin{aligned}|f(z)|\geqslant C_{n_k}=C'_{n_k}-\frac{g}{n_k}&>|f(z_{n_k})|-\frac{g}{n_k}\\&=|f(z_0)|+|f(z_{n_k})|-|f(z_0)|-\frac{g}{n_k}.\end{aligned}$$

令 $k\to\infty$,我们有

$$\lim_{k\to\infty}\frac{g}{n_k}=0,\quad \lim_{k\to\infty}\left(|f(z_{n_k})|-|f(z_0)|\right)=0.$$

这样一来,我们得到,$|f(z)|\geqslant|f(z_0)|$.这就完成了定理的证明.

代数基本定理的证明

由(1)式,我们知道,$f(0)=a_n$.根据引理 2,总存在一个正数 R,当 $|z|>R$ 时,$|f(z)|>|a_n|$.从而我们可将注意力集中于考虑圆盘 $|z|\leqslant R$.由定理 1,$|f(z)|$ 在圆盘 $|z|\leqslant R$ 的某一点 z_0 处取得最小值:$|f(z_0)|\leqslant|f(0)|$.若 $|f(z_0)|=0$,则也有 $f(z_0)=0$,从而定理得证.

今假定 $|f(z_0)|\neq 0$.我们来证明,$|f(z_0)|\neq 0$ 一定会引出矛盾.

考虑多项式 $Q(z) = f(z+z_0)/f(z_0)$. $Q(z)$ 是多项式 $f(z+z_0)$ 除以常数 $f(z_0)$，所以它仍是一个 n 次多项式. 注意到

$$|Q(z)| = \frac{|f(z+z_0)|}{|f(z_0)|} \geqslant \frac{|f(z_0)|}{|f(z_0)|} = 1, \text{而} |Q(0)| = 1,$$

所以 $|Q(z)|$ 在 $z=0$ 有最小值：$|Q(0)| = 1$. 易见，$Q(0) = 1$，所以 $Q(z)$ 可以表示为

$$Q(z) = 1 + az^m + z^{m+1}R(z)(m \geqslant 1),$$

这里 $a \neq 0$，az^m 是 $Q(z)$ 中系数不为 0 的次数最低的那一项，$z^{m+1}R(z)$ 表示高次项，$R(z)$ 是复系数多项式. 我们在讲复数时曾讨论过方程

$$ay^m + 1 = 0 \Leftrightarrow y^m + \frac{1}{a} = 0$$

的根的求解的问题. 设 c 是它的一个根：$ac^m + 1 = 0$，或 $ac^m = -1$. 这时，

$$\begin{aligned} Q(cz) &= 1 + ac^m z^m + c^{m+1}z^{m+1}R(cz) \\ &= 1 - z^m + c^{m+1}z^{m+1}R(cz). \end{aligned} \tag{2}$$

$|Q(cz)|$ 在 $z=0$ 有最小值，最小值为 1.

我们从 $|f(z_0)| \neq 0$ 引出 $|Q(0)| = 1$ 是 $|Q(z)|$ 的最小值. 现在我们来证明 1 不是 $|Q(z)|$ 的最小值，从而得到一个矛盾. 这个矛盾指出，假定 $|f(z_0)| \neq 0$ 是错误的，由此得到代数基本定理. 下面我们来证明 1 不是 $|Q(z)|$ 的最小值.

因为多项式 $R(cz)$ 在圆盘 $|z| \leqslant R$ 上是有界的，所以可以找到一个正数 T，使得 $|R(cz)| \leqslant T$. 根据(2)，

$$\begin{aligned} |Q(cz)| &\leqslant |1 - z^m| + |c|^{m+1}|z|^{m+1}|R(cz)| \\ &\leqslant |1 - z^m| + |c|^{m+1}|z|^{m+1}T. \end{aligned}$$

令 z 沿正实轴趋于 0，这时 $|1 - z^m| = 1 - z^m$，于是

$$\begin{aligned} |Q(cz)| &\leqslant 1 - z^m + |c|^{m+1}z^{m+1}T \\ &= 1 - z^m(1 - z|c|^{m+1}T) \leqslant 1 - \frac{1}{2}z^m. \end{aligned}$$

这是因为当 z 充分小时，$1 - z|c|^{m+1}T \geqslant \frac{1}{2}$. 上式指出，当 z 充分小时，

$$|Q(cz)| < 1 = Q(0).$$

这与 $|Q(0)|$ 取得最小值是矛盾的. 这个矛盾建立了代数基本定理.

代数基本定理的证明是存在性的证明. 它既没有告诉我们根的何处，也没有告诉我们如何去计算这个根.

几何解释

我们对代数基本定理作一几何考察. 在复平面的每一个点 z 上安置一个立坐标 t，它的长等于多项式 $f(z)$ 在这一点的模 $|f(z)|$. 这些立坐标的端点形成一个曲面，我们把它叫作多项式 $f(z)$ 的模曲面. 因为 $|f(z)| \geqslant 0$，所以这个模曲面无论在什么地方都不会降到复平面的下面. 同时我们还注意到，复平面上的任一点 z，曲面上有且只有一点与它对应. 也就是说，整个曲面在复平面上只有一叶. 由于 $|f(z)|$ 是 z 的连续函数，所以当 z 在复平面上变动时，这个曲面的立坐标连续地变动. 代数基本定理指出，多项式 $f(z)$ 的模曲面至少有一点接触到复平面. 事实上，在 $|f(z)|$ 的每一极小值处都给出 $f(z)$ 的一个根(图 10－1). 极小值的个数等于多项式的不同根的个数. 这些极小值在复平面上把多项式的模曲面支撑了起来.

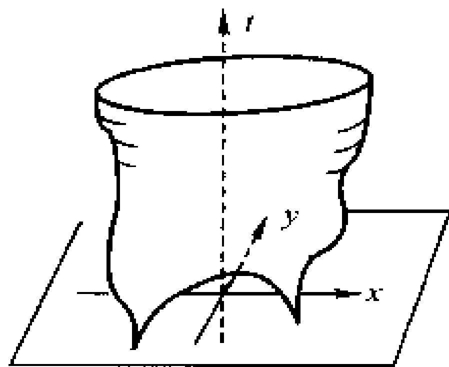


图 10－1

§10.3 多项式的根的分布问题

在许多实际应用中都需要研究多项式的根的分布问题. 问题是，通过多项式的系数，我们能不能得到关于多项式的根的分布的信息，特别是关于多项式的实根的个数的信息. 我们就来着手研究这些

问题.

10.3.1　多项式的单根和重根

前面曾指出，若 a 是多项式 $f(x)$ 的根，则 $(x-a)$ 可以整除 $f(x)$. 如果 $(x-a)$ 可以整除 $f(x)$，而 $(x-a)^2$ 不能整除 $f(x)$，则 a 称为多项式 $f(x)$ 的单根. 如果 $(x-a)^k$ 可以整除 $f(x)$，而 $(x-a)^{k-1}$ 不能整除 $f(x)$，则 a 称为多项式 $f(x)$ 的 k 重根.

k 重根 a 常常看成 k 个相等的根. 我们约定，每个根所算的次数就是它的重数. 这样 n 次多项式就有 n 个根.

例　$f(x)=x(x-1)^2(x-2)^3$ 是一个 6 次多项式. 它有一个单根 $x=0$，一个 2 重根 $x=1$ 和一个 3 重根 $x=2$.

我们先研究多项式的根和它的导数的根的关系. 这要用一些微分学的知识. 如所周知，若

$$f(z)=z^n+a_1z^{n-1}+a_2z^{n-2}+\cdots+a_{n-1}z+a_n,$$

则它的导数是

$$f'(z)=nz^{n-1}+(n-1)a_1z^{n-2}+\cdots+a_{n-1}.$$

在微分学中，我们只研究实变量的函数的导数，现在容许变量为复变量；导数的定义仍然是函数的增量与自变量的增量之比的极限.

定理 1　a) 多项式的单根不是它的导数的根；b) 多项式的重根是它的导数的根，但重数少 1.

证　只需证 b)，因为 a) 蕴含在 b) 中. 设

$$f(x)=(x-a)^kf_1(x)(1\leqslant k\leqslant n),$$

其中 $f_1(x)$ 不能被 $x-a$ 整除，即 $f_1(a)\neq 0$. 那么

$$\begin{aligned}f'(x)&=k(x-a)^{k-1}f_1(x)+(x-a)^kf_1'(x)\\&=(x-1)^{k-1}[kf_1(x)+(x-a)f_1'(x)]\\&=(x-1)^{k-1}F(x),\end{aligned}$$

其中 $F(x)=kf_1(x)+(x-a)f_1'(x)$，它不能被 $x-a$ 整除，因为 $F(a)=kf_1(a)\neq 0$.

因此,当 $k=1$ 时,$f'(x)$ 不能被 $x-a$ 整除.当 $k>1$ 时,$f'(x)$ 可被 $(x-a)^{k-1}$ 整除,但不能被 $(x-a)^k$ 整除.定理得证.

10.3.2 罗尔定理和它的推论

定理 2(罗尔) 若函数 $f(x)$ 在闭区间 $[a,b]$ 上连续,在开区间 (a,b) 上可微,并且 $f(a)=f(b)$,则存在一点 $c\in(a,b)$,使得 $f'(c)=0$.

有了罗尔定理我们可以得到下面的有趣推论.

系 1 如果 n 次多项式 $f(x)$ 的一切根都是实的,则它的导数的根也是实的,并且在 $f(x)$ 的相邻两个根之间有 $f'(x)$ 的一个根,而且是实根.

证 设 $f(x)$ 有 k 个不同的根,$x_1<x_2<\cdots<x_k$.它们的重数分别是 $m_1,m_2,\cdots,m_k$.从而,

$$m_1+m_2+\cdots+m_k=n.$$

根据定理 2,导数 $f'(x)$ 有根 $x_1,x_2,\cdots,x_k$,它们的重数分别是

$$m_1-1,m_2-1,\cdots,m_k-1.$$

由罗尔定理,$f'(x)$ 在区间 $(x_1,x_2),(x_2,x_3),\cdots,(x_{k-1},x_k)$ 里还有根 $y_1,y_2,\cdots,y_{k-1}$.这样一来,$f'(x)$ 的实根的个数至少是

$$(m_1-1)+(m_2-1)+\cdots+(m_k-1)+k-1=n-1.$$

但 $f'(x)$ 是 $n-1$ 次多项式,它有 $n-1$ 个根,因此 $f'(x)$ 的一切根都是实的,且 $y_1,y_2,\cdots,y_{k-1}$ 是单根.除了 $y_1,y_2,\cdots,y_{k-1},x_1,x_2,\cdots,x_k$ 外 $f'(x)$ 没有其它根.

系 2 如果 n 次多项式 $f(x)$ 的一切根都是实的,并且其中有 p 个根是正根(计算根的重数),则 $f'(x)$ 有 p 个正根,或 $p-1$ 个正根.

证 设 $x_1<x_2<\cdots<x_k$ 是 $f(x)$ 的正根,它们的重数分别是 $m_1,m_2,\cdots,m_k$,则

$$m_1+m_2+\cdots+m_k=p.$$

导数 $f'(x)$ 将有下列正根:$x_1,x_2,\cdots,x_k$,它们的重数分别是

$$m_1-1, m_2-1, \cdots, m_k-1.$$

$f'(x)$ 在区间 $(x_1,x_2),(x_2,x_3),\cdots,(x_{k-1},x_k)$ 里还有单根 $y_1,y_2,\cdots,y_{k-1}$. 可能还有一个单根 y_0 位于 (x_0,x_1) 之间,这里 x_0 是 $f(x)$ 的最大的非正根. 因此 $f'(x)$ 的正根数等于

$$(m_1-1)+(m_2-1)+\cdots+(m_k-1)+k-1=p-1,$$

或

$$(m_1-1)+(m_2-1)+\cdots+(m_k-1)+(k-1)+1=p.$$

证毕.

10.3.3　笛卡儿符号定则

笛卡儿在 1637 年发表了名著《几何学》,在这部书中他第一次阐述了解析几何,这就是解析几何的诞生. 同时他还顺便给出了一个著名的代数定理,这就是笛卡儿符号定则. 如果知道多项式的根都是实的,利用这个定理可以根容易地求出它的正根的个数. 这是一个很漂亮的结果.

为此需要引进一个多项式的系数序列的变号数的概念. 设

$$f(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n.$$

假定 $a_0>0$,并写出序列

$$a_0,a_1,a_2,\cdots,a_{n-1},a_n.$$

去掉其中等于 0 的系数. 考察序列中一切相邻的数对. 如果这样的数对中,两数的符号不同,那么就叫作一个变号. 变号数的总和叫做一个多项式的系数序列的变号数.

例　多项式

$$x^4+2x^3-13x^2-14x+24$$

的系数序列是

$$1,2,-13,-14,24.$$

这个序列有 2 次变号,因而该多项式的变号数是 2.

例　多项式

$$x^7+3x^5-5x^4-8x^3+7x+2$$

的系数的序列是

$$1,3,-5,-8,7,2.$$

这个序列也有 2 次变号，因而该多项式的变号数也是 2.

定理 3(笛卡儿符号定则) 如果多项式

$$f(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n$$

的系数是实数，并且它的一切根也都是实数，那么它的正根的个数(计算重数)等于它的系数序列的变号数.

引理 若 $f(x)$ 有 p 个正根(计算重数)，则 $(-1)^p$ 是 $f(x)$ 的最后一个不等于 0 的系数的符号.

证 事实上，设多项式的最后一个不为 0 的系数是 a_k，则

$$f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_kx^{n-k}=a_0x^{n-k}\left(x^k+\cdots+\frac{a_k}{a_0}\right)$$

$$=a_0x^{n-k}(x-x_1)\cdots(x-x_p)(x-x_{p+1})\cdots(x-x_k).$$

式中 $x_1,\cdots,x_p$ 是 $f(x)$ 的正根，$x_{p+1},\cdots,x_k$ 是 $f(x)$ 的负根，每一个根所算的次数就是它的重数. 根据韦达定理，

$$a_k=a_0(-1)^px_1\cdots x_p(-x_{p+1})\cdots(-x_k).$$

由于 $a_0,x_1,\cdots,x_p,-x_{p+1},\cdots,-x_k$ 都是正数，所以 a_k 的符号是 $(-1)^p$.

定理 3 的证明 用归纳法证明. 对多项式的次数 n 作归纳法.

对于一次多项式定理是显然的. 事实上，

$$a_0x+a_1=0\Rightarrow x=-\frac{a_1}{a_0}.$$

只有 a_0,a_1 的符号相反时根才是正的.

现在假设定理对实系数的一切 $n-1$ 次多项式已经证明. 我们来证明定理对 n 次多项式也成立. 设

$$f(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n.$$

今分为两种情况讨论：

1. $a_n = 0$. 这时考虑多项式

$$f_1(x) = a_0x^{n-1} + a_1x^{n-2} + \cdots + a_{n-1}.$$

多项式 $f(x)$ 与 $f_1(x)$ 的正根是相同的，它们的系数序列的变号数也是相同的. 对于多项式 $f_1(x)$ 而言，笛卡儿法则是正确的，因此对多项式 $f(x)$ 也是正确的.

2. $a_n \neq 0$. 我们来考察导数

$$f'(x) = na_0x^{n-1} + (n-1)a_1x^{n-2} + \cdots + a_{n-1}.$$

现在看 a_n 与导数的最后一个不等于 0 的系数. 当它们的符号相同时，$f(x)$ 的变号数与 $f'(x)$ 的变号数相同；当它们的符号相异时，$f(x)$ 的变号数多一个.

由引理知，在第一种情形下 $f(x)$ 与 $f'(x)$ 的正根数有相同的奇偶性，而在第二种情形下则相反. 由系 2，$f(x)$ 与 $f'(x)$ 的正根数相等，或者 $f(x)$ 的正根数多一个. 利用这一点，我们断言在第一种情况下，$f(x)$ 与 $f'(x)$ 的正根数一样多，在第二种情况下 $f(x)$ 的正根数就多一个. 由归纳法假设，笛卡儿定则对于 $f'(x)$ 是正确的，即 $f'(x)$ 的正根数等于它的变号数，因此在两种情形下 $f(x)$ 正根的个数(计算重数) 都等于它的系数序列的变号数. 证毕.

注 1 这个定理很重要. 在许多实际问题中常常知道方程的一切根都是实的. 有了这个定理我们就可立即知道，方程有多少个正根多少个负根和多少个零根.

解 方程

$$x^3 - 7x + 6 = 0$$

的 3 个根都是实数：1，2，－3. 其系数的变号数是 2. 它有 2 个正根一个负根.

注 2 在多项式 $f(x)$ 中令 $x = y + a$，这里 a 是任意给定的数，并写出多项式 $f(y+a)$. 易见，x_i 是 $f(x)$ 的根，当且仅当 $y_i = x_i - a$ 是 $f(y+a)$ 的根. 即

$$f(x_i)=0 \Leftrightarrow f(y_i+a)=0.$$

并且
$$y_i>0 \Rightarrow x_i>a,$$
即若 y_i 是 $f(y+a)$ 的正根，则 x_i 是多项式 $f(x)$ 的大于 a 的根.

例　求方程 $x^3-7x+6=0$ 的 $x>3$ 的根的个数.

解　令 $x=y+3$，代入方程，得
$$(y+3)^3-7(y+3)+6=0.$$
展开并化简，得
$$y^3+9y^2+20y+12=0.$$

这个方程的变号数是0，没有正根. 因而，原方程没有 $x>3$ 的根.

现在设多项式 $f(x)$ 的一切根都是实数. 问，它在 a 与 $b(b>a)$ 之间的根的个数是多少？前面的论述告诉我们，

多项式 $f(x+a)$ 的变号数 $=$ $f(x)$ 的大于 a 的根的个数

多项式 $f(x+b)$ 的变号数 $=$ $f(x)$ 的大于 b 的根的个数

所以 $f(x)$ 在 a,b 间的根的个数等于 $f(x+a)$ 的变号数减去 $f(x+b)$ 的变号数.

例　求方程 $x^3-7x+6=0$ 在区间 $(0,3)$ 内的根的个数.

解　已知，$x>3$ 时，方程的根的个数是0，$x>0$ 时，方程的根的个数是2，由此知，方程在 $(0,3)$ 内的个数是2.

10.3.4　辐角原理

下面给出的辐角原理可以用来研究多项式的根的分布问题.

设给定一个多项式 $f(z)$ 和复平面上的一个区域 D，我们来讨论在这个区域内 $f(z)$ 有多少根. 我们不再限定多项式的系数都是实系数；它们可以是实的或复的. 假定区域 D 是由一条封闭的曲线围成的(图10－2)，并且在 D 的边界上没有 $f(z)$ 的根.

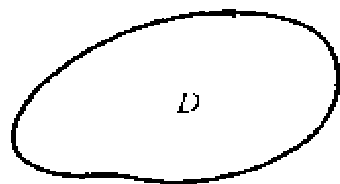


图10－2

考虑两个辅助平面，一个是区域 D 所在的平面，称为 z 平面；另一个平面是 w 平面，$w = f(z)$ 是这个平面上的点. 当 z 在 z 平面上变化时，多项式的值 $w = f(z)$ 就在 w 平面上变化.

D 的边界有两个方向. 其中一个方向是这样的：沿着这个方向走，区域 D 总在它的左边，这个方向称为边界的正方向. 另一个方向称为它的反方向.

设想点 z 沿着 D 的边界的正方向走一周，相应地，$w = f(z)$ 就在 w 平面上画出一条封闭的曲线(图 10 − 3). 根据假设，$f(z)$ 在 D 的边界上没有 0 点，所以这条曲线不过原点. 下面的定理给出了 $f(z)$ 在 D 内的根的个数.

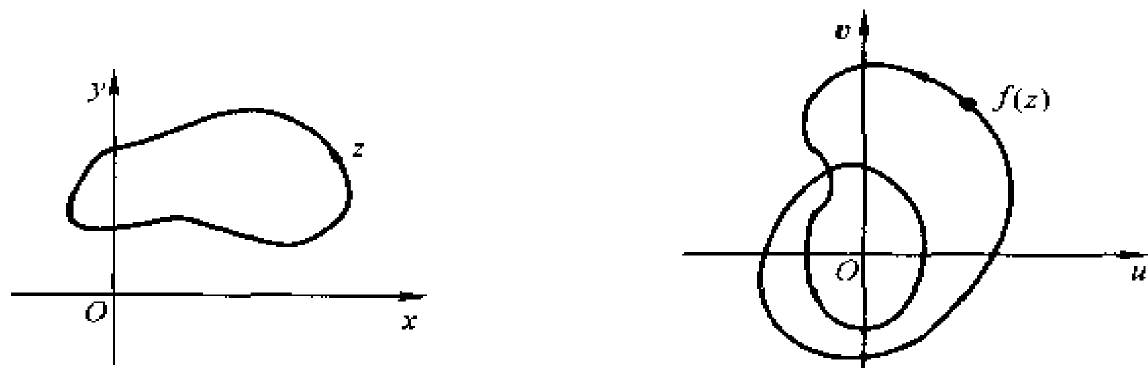


图 10 − 3

定理 4　设区域 D 由一条闭曲线 C 所围成，并且多项式

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n$$

在 C 上没有 0 点，那么 $f(z)$ 在 D 内根的个数等于当 z 沿 C 的正向通过一次时 $w = f(z)$ 在 w 平面上绕原点的圈数.

证　根据代数基本定理，$f(z)$ 可以分解为一次因子的连乘积

$$f(z) = a_0(z - z_1)(z - z_2)\cdots(z - z_n),$$

这里 $z_1, z_2, \cdots, z_n$ 是多项式 $f(z)$ 的 n 个根. 因为复数乘积的辐角等于各因子的辐角的和，所以

$$\arg f(z) = \arg a_0 + \arg(z - z_1) + \arg(z - z_2) + \cdots + \arg(z - z_n),$$

用 $\Delta\arg f(z)$ 表示 z 绕 C 的正方向环行一周时 $f(z)$ 的辐角的改变量. 易见,这个量是 2π 的整数倍. 用 $\Delta\arg(z - z_i)$ 表示 z 绕 C 的正方向环行一周时 $(z - z_i)$ 的辐角的改变量. 于是我们有下面的关系式.

$$\Delta\arg f(z) = \Delta\arg a_0 + \Delta\arg(z - z_1) + \Delta\arg(z - z_2) + \cdots + \Delta\arg(z - z_n).$$

因为 a_0 是一个常数,其辐角不会改变,所以 $\Delta\arg a_0 = 0$. $z - z_1$ 可以用从点 z_1 到 z 的向量来表示. 若 z_1 在 D 的内部,那么在几何上看,当 z 沿 C 绕行一周时,向量 $z - z_1$ 以 z_1 为中心绕过一整周(图 10 - 4). 因此 $\Delta\arg(z - z_1) = 2\pi$. 现在假定 z_2 位于区域 D 之外. 在这种情况下,当 z 沿 C 绕行一周时,向量 $z - z_2$ 没有绕过 z_2,因此 $\Delta\arg(z - z_2) = 0$. 我们可以用这种方法考察 $f(z)$ 的所有根. 由此我们得到结论: $\Delta\arg f(z)$ 等于 2π 乘以 $f(z)$ 在区域 D 内的根的个数. 因此 $f(z)$ 在区域 D 内的根的个数等于点 $w = f(z)$ 在 w 平面上绕原点的圈数. 证毕.

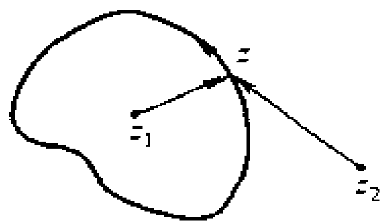


图 10 - 4

通常我们并不直接用辐角原理来计算某区域内一个多项式的根的个数,而借助下面的路西定理来计算.

定理 5(路西定理)　设 $P(z)$ 和 $Q(z)$ 是两个多项式,C 是一条闭曲线. 若在 C 上,$P(z)$ 和 $Q(z)$ 满足 $|P(z)| > |Q(z)|$,则在 C 的内部 $P(z) + Q(z)$ 和 $P(z)$ 有相同的零点个数.

证　我们利用辐角原理来求 $P(z) + Q(z)$ 的零点个数. 将 $P(z) + Q(z)$ 改写为

$$P(z) + Q(z) = P(z)\left\{1 + \frac{Q(z)}{P(z)}\right\}.$$

注意,在曲线 C 上,$|P(z)| > |Q(z)|$,所以在曲线 C 上 $P(z)$ 不会为零. 于是

$$\arg[P(z)+Q(z)]=\arg P(z)+\arg\left\{1+\frac{Q(z)}{P(z)}\right\}.$$

由于$\left|\frac{Q(z)}{P(z)}\right|<1$,所以向量$1+\frac{Q(z)}{P(z)}$的终点画出一条闭曲线,这条闭曲线在以1为中心,以1为半径的圆内.因此这个向量没有绕过原点.这样一来,当z绕行C一周后,$\arg\left\{1+\frac{Q(z)}{P(z)}\right\}$的值没有改变.所以

$$\Delta\arg[P(z)+Q(z)]=\Delta\arg P(z).$$

由辐角原理推出,$P(z)+Q(z)$和$P(z)$在C内有相同的零点个数.证毕.

例 求在圆$|z|<1$内方程$z^8-4z^5+z^2-1=0$的根的个数.

解 我们利用路西定理.先将方程表示成$P(z)+Q(z)$的形式.取

$$P(z)=-4z^5\quad Q(z)=z^8+z^2-1,$$

在圆周$|z|=1$上比较$|P(z)|$和$|Q(z)|$,我们有

$$|Q(z)|=|z^8+z^2-1|\leqslant|z^8|+|z^2|+1=3,$$

$$|P(z)|=|4z^5|=4.$$

从而,$|Q(z)|<|P(z)|(|z|=1)$.根据路西定理,$P(z)+Q(z)=z^8-4z^5+z^2-1$在圆$|z|<1$内零点的个数与$P(z)=-4z^5$的零点个数相同.$P(z)$在$|z|<1$内的零点个数是5,所以原方程在$|z|<1$内的零点个数也是5.

§10.4 实根的近似计算法

实根的近似计算在应用中具有重要的意义,计算机的广泛使用使得近似计算更加重要.本节介绍三种求多项式的根的近似值的方法.连续函数的中间值定理具有基础的地位,我们先给出这一定理.

定理1 若$f(x)$在区间$[a,b]$上连续,且$f(a)<0,f(b)>0$,

则一定在区间$[a,b]$上至少存在一点c,使得$f(c)=0$.(证明见微积分部分).

设

$$f(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n.$$

显然,$f(x)$是实轴上的连续函数.

10.4.1 二分法

如果在区间$[a,b]$上$f(a)$与$f(b)$异号,那么$f(x)$在$[a,b]$上一定有一个根.为叙述方便计,我们假定$f(a)<0,f(b)>0$,且$f(x)$在$[a,b]$上只有一个根α.我们用二分法来确定α的近似值.

取
$$x_1=\frac{a+b}{2}.$$

如果$f(x_1)=0$,那么x_1就是$f(x)$的根.如果$f(x_1)\neq 0$,那么$f(x_1)$必和$f(a)$与$f(b)$中之一异号.不妨设$f(x_1)$与$f(b)$异号(图 10 - 5).在区间内取

$$x_2=\frac{x_1+b}{2}.$$

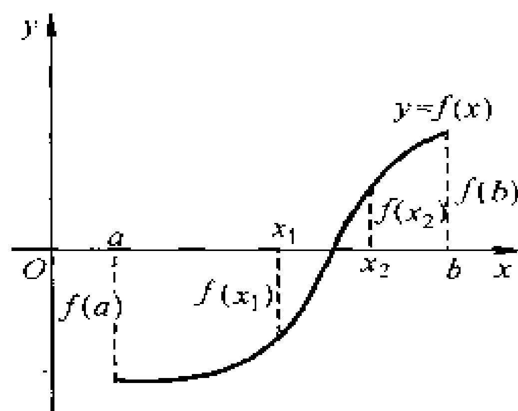


图 10 - 5

如果$f(x_2)=0$,那么x_2就是$f(x)$的根.如果$f(x_2)\neq 0$,那么$f(x_2)$必和$f(x_1)$与$f(b)$中之一异号.不妨设$f(x_2)$与$f(x_1)$异

号.在区间(x_1,x_2)内取

$$x_3=\frac{x_1+x_2}{2}.$$

如此继续下去,就可得到任意指定的精确度的α的近似值.

例 已知方程$f(x)=2x^3+6x-5=0$在$[0,1]$内有一个实根α.求一个精确到0.01的近似根.

解 容易算出,$f(0)=-5<0$,$f(1)=3>0$.取$x_1=0.5$.由于$f(0.5)<0$,所以α落在区间$(0.5,1)$内.取$x_2=\frac{0.5+1}{2}=0.75$.

由于$f(0.75)>0$,所以α落在区间$(0.5,0.75)$内.取$x_3=\frac{0.5+0.75}{2}=0.625$,而$f(0.625)<0$.于是$\alpha$落在区间$(0.625,0.75)$内.取$x_4=\frac{0.625+0.75}{2}=0.688^-$.$f(0.688)<0$.所以$\alpha$落在区间$(0.688,0.75)$内.取$x_5=\frac{0.688+0.75}{2}=0.719$.$f(0.719)>0$.所以$\alpha$落在区间$(0.688,0.719)$内.$f(0.704)<0$取$x_6=\frac{0.688+0.719}{2}$.$\alpha$落在区间$(0.704,0.719)$内.我们的目的是取精确度0.01的近似值.从上面的计算可知,取$\alpha\approx0.7$,或取$\alpha\approx0.71$.

10.4.2 插值法

另一种方法是插值法.这种方法的基本思想是,利用弦与x轴的交点来求根的近似值.如图10-6,用直线连接点$A(a,f(a))$和点$B(b,f(b))$.AB与x轴交于点$(x_1,0)$.如果$f(x_1)=0$,那么x_1就是$f(x)$的根.如果$f(x_1)\neq0$,那么可以把x_1看作α的一个近似值.下面给出求x_1的方法.弦AB的方程是

$$\frac{y-f(a)}{f(b)-f(a)}=\frac{x-a}{b-a}\Leftrightarrow x-a=\frac{(y-f(a))(b-a)}{f(b)-f(a)}.$$

把$(x_1,0)$代入上式,得

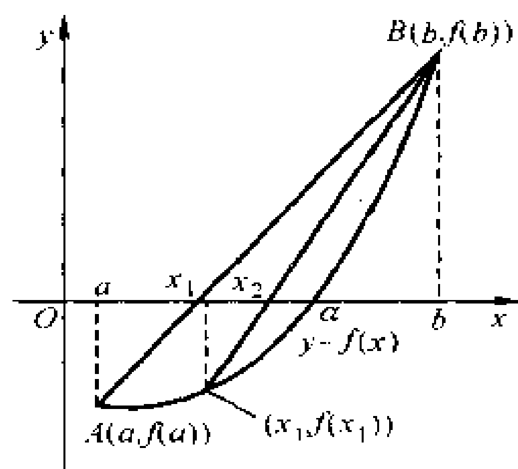


图 10－6

$$x_1 = a - \frac{f(a)(b-a)}{f(b)-f(a)},$$

即
$$x_1 = \frac{af(b)-bf(a)}{f(b)-f(a)}. \tag{1}$$

这就是求 x_1 的公式.

如果 $f(x_1)$ 与 $f(b)$ 异号,可用直线连接点 $(x_1, f(x_1))$ 与 $(b, f(b))$,这条直线与 x 轴交于点 $(x_2, 0)$. 如果 $f(x_2) = 0$,那么 x_2 就是 $f(x)$ 的根. 如果 $f(x_2) \neq 0$,那么可以把 x_2 看作 α 的一个近似值. x_2 的值可用公式(1)求出:

$$x_2 = \frac{bf(x_1)-x_1f(b)}{f(x_1)-f(b)}.$$

如此继续下去就可以得到任意指定的精确度的 α 的近似值.

插值法比二分法逼近实根的速度要快一些,因此计算步骤也少一些.

例 已知方程 $f(x) = x^3 - 2x - 5 = 0$ 在 $[2,3]$ 内有一个实根 α. 求一个精确到 0.001 的近似根.

解 用插值法计算. 已知 $a = 2, b = 3, f(2) = -1 < 0, f(3)$

$= 16 > 0$. 由公式(1)得，

$$x_1 = \frac{af(b) - bf(a)}{f(b) - f(a)} = \frac{2f(3) - 3f(2)}{f(3) - f(2)}$$
$$= \frac{2 \cdot 16 - 3(-1)}{16 - (-1)} \approx 2.1.$$

因为 $f(2.1) = 0.061 > 0$，它与 $f(2) < 0$ 异号. 因此 α 在区间(2, 2.1)内. 由公式(1)得，

$$x_2 = \frac{2f(2.1) - 2.1f(2)}{f(2.1) - f(2)} = \frac{2 \times 0.061 - 2.1(-1)}{0.061 - (-1)} \approx 2.09.$$

因为 $f(2.09) = -0.015 < 0$，它与 $f(2.1) > 0$ 异号，因此 α 在区间(2.09,2.1)内. 由公式(1)得，

$$x_3 = \frac{2.09f(2.1) - 2.1f(2.09)}{f(2.1) - f(2.09)}$$
$$= \frac{2.09 \times 0.061 - 2.1(-0.051)}{0.061 - (-0.051)} \approx 2.098.$$

所求根 α 的小数点后三位的的近似值是 2.098.

10.4.3 牛顿法

插值法的基本思想是借助曲线 $y = f(x)$ 的割线与 x 轴的交点求 $f(x)$ 的根的近似值. 牛顿法的基本思想是用切线代替割线. 为此，除了前面的假定外，我们再假定在区间 $[a, b]$ 上 $f'(x) \neq 0$. 如图 10－7 所示，过点 $A(a, f(a))$ 作曲线 $y = f(x)$ 的切线 AT，AT 交 x 轴于点 $(x_1, 0)$. 如果 $f(x_1) = 0$，那么 x_1 就是 $f(x)$ 的根. 如果 $f(x_1) \neq 0$，那么可以把 x_1 看作 α 的一个近似值. 下面给出求 x_1 的方法.

切线 AT 的方程是

$$y - f(a) = f'(a)(x - a).$$

把 $(x_1, 0)$ 代入上式得

$$x_1 = a - \frac{f(a)}{f'(a)}. \tag{2}$$

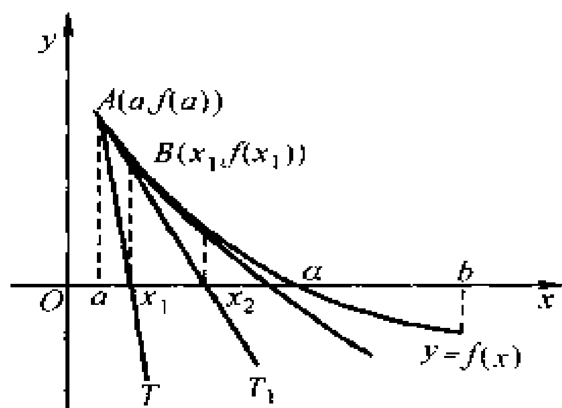


图 10 - 7

如果 $f(x_1)$ 与 $f(b)$ 异号，可以过点 $B(x_1,f(x_1))$ 作曲线的切线 BT_1，BT_1 交 x 轴交于点$(x_2,0)$. 如果 $f(x_2)=0$，那么 x_2 就是 $f(x)$ 的根. 如果 $f(x_2)\neq 0$，那么可以把 x_2 看作 α 的一个近似值. x_2 的值可用公式(2) 求出：

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

如此继续下去就可以得到任意指定的精确度的 α 的近似值.

例　已知方程 $f(x)=x^3-x-9=0$ 在$[2,3]$内有一个实根 α. 求一个精确到 0.001 的近似根.

解　已知 $a=2, b=3, f(2)=-3<0, f(3)=15>0$. 在区间$[2,3]$上，$f'(x)=3x^2-1>0$. 今用牛顿法求 α 的近似值. 根据公式(2)，

$$x_1 = a - \frac{f(a)}{f'(a)} = 2 - \frac{-3}{11} \approx 2.27.$$

因为 $f(2.27)=0.427>0$，它与 $f(2)$ 异号，所以 α 在区间$(2,2.27)$内. 由公式(2) 得，

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 2.27 - \frac{0.427}{14.458} \approx 2.240.$$

所求根 α 的小数点后三位的的近似值是 2.240.

习　题

求下列方程在指定区间内的实根的近似值:

1. $x^3+x-3=0$,在(1,2)内;(用平分法)
2. $x^3+2x-20=0$,在区间(2,3)内;(用插值法)
3. $x^3-6x+\frac{\pi}{10}=0$,在区间(0,1)内;(用牛顿法)
4. $x^3+6x^2+10x-2=0$,在区间(0,1)内;(用平分法)
5. $x^3+x^2-2x-1=0$,在区间(-1,0)内;(用插值法)
6. $x^4-10x^2-4x+8=0$,在区间(3,4)内.(用牛顿法)
7. 计算下列各方程在单位圆 $|z|<1$ 内的根的个数:

1) $2z^5-z^3+3z^2-z+8=0$,

2) $z^7-5z^4+z^2-2=0$,

3) $z^9-2z^6+z^2-8z-2=0$.

第十一章　双曲几何的庞加莱模型

对自然界的深刻研究是数学最富饶的源泉.

Joreph Fourion

数学的无穷无尽的诱人之处在于,它的最棘手的悖论能够盛开出美丽的理论之花.

非欧几何证明了数学是人的手工作品,它仅仅受到思维规律所规定的限制.

E. Kasner and J. Newman

小说家发明人物、对话和情节,关于它们,他既是作家又是主人;数学家随心所欲地设计公设,使它的数学体系奠基于其上;二者十分相像.小说家和数学家在选择和处理他们的素材时,都可能受他们所处环境的限制;但是没有什么超人的、永恒的必然性迫使他们去创造某人物或发明某体系.

E. T. Bell

我们曾经介绍了欧几里得的第五公设和非欧几何的诞生史,这一讲我们来扼要介绍双曲几何的庞加莱模型.我们知道,平行公设的研究导致了非欧几何的诞生.这件事无论在数学史上,还是在科学史上都具有划时代的意义.它彻底改变了人们对数学本身是什么的认识,也改变了人们对空间的认识.非欧几何又为日后相对论的诞生奠定了基础.作为一个有一定文化修养的人应该对非欧几何有所了解.同时,只有学点非欧几何,才会对欧氏几何有更全面、更正确和更深入的认识.

读者从中学时代起就已熟悉了欧氏几何的第五公设.或许已经

听说过,存在一种非欧几何,在这种几何中,过直线外一点可以作无穷多条直线与所给直线不相交.在最初听到有这种几何时,你或许会惊讶,不可理解,甚至觉得它违反常识.这可能吗?在没有真正见到这种几何时,这种神秘感或许永远无法驱除.现在我们就来着手给出这种几何的一个模型.所需要的预备知识很少,只要知道中学的复数知识和一点空间解析几何的知识就够了.

§11.1 球极平面投影

球极平面投影在复分析中的重要作用在于,为复平面引进了一个无穷远点,使复平面变成一个紧曲面,成为处理亚纯函数的有力工具.这里借助球极平面投影把直线与圆作统一处理.

11.1.1　直线与圆的复数形式

我们先给出直线和圆的复数表示,这对后面的内容十分有用.从平面解析几何中,我们知道在取定坐标系后,一条直线的方程是

$$ax + by + c = 0, a, b, c \in \mathbf{R}. \tag{1}$$

若令 $z = x + \mathrm{i}y$,则 $\overline{z} = x - \mathrm{i}y$,从而

$$x = \frac{1}{2}(z + \overline{z}), y = \frac{1}{2i}(z - \overline{z}).$$

把它们代入(1)并化简,得到

$$(a - \mathrm{i}b)z + (a + \mathrm{i}b)\overline{z} + 2c = 0.$$

令 $\beta = a + \mathrm{i}b$,则上面的方程可化为

$$\overline{\beta}z + \beta\overline{z} + 2c = 0. \tag{2}$$

其中 $\beta \neq 0$ 是复数,而 c 是实数,这就是直线方程的复数形式.

下面研究圆的方程.设 z_0 是复平面 C 上的任意一点,以 z_0 为中心,以 r 为半径的圆的方程具有形式

$$|z - z_0| = r. \tag{3}$$

此式等价于

$$|z-z_0|^2=\overline{(z-z_0)}(z-z_0)=r^2.$$

展开即得

$$z\bar{z}-\bar{z}_0z-z_0\bar{z}+|z_0|^2-r^2=0.$$

两边乘以实数 $\alpha\neq 0$,得

$$\alpha z\bar{z}-\alpha\bar{z}_0z-\alpha z_0\bar{z}+\alpha(|z_0|^2-r^2)=0.$$

令 $\beta=-\alpha z_0,\gamma=\alpha(|z_0|^2-r^2)$,上式就化为

$$\alpha z\bar{z}+\beta\bar{z}+\bar{\beta}+\gamma=0. \tag{4}$$

其中 α,γ 均为实数,且满足条件

$$\alpha\neq 0,\alpha\gamma<|\beta|^2. \tag{5}$$

事实上,

$$|\beta|^2=\alpha^2|z_0|^2,\alpha\gamma=\alpha^2|z_0|^2-\alpha^2r^2=|\beta|^2-\alpha^2r^2.$$

从而

$$\alpha\gamma<|\beta|^2.$$

反过来,满足条件(5)的方程(4)表示一个圆周.事实上(4)可化为

$$\left(z+\frac{\beta}{\alpha}\right)\left(\bar{z}+\frac{\bar{\beta}}{\alpha}\right)+\frac{\gamma}{\alpha}-\frac{|\beta|^2}{\alpha^2}=0.$$

或

$$\left|z+\frac{\beta}{\alpha}\right|^2=\frac{|\beta|^2}{\alpha^2}-\frac{\gamma}{\alpha}. \tag{6}$$

条件(5)指出 $\frac{|\beta|^2}{\alpha^2}-\frac{\gamma}{\alpha}>0$.因而(6)表示一个以 $-\frac{\beta}{\alpha}$ 为中心,以 $\sqrt{\frac{|\beta|^2}{\alpha^2}-\frac{\gamma}{\alpha}}$ 为半径的圆.

在(4)中,当 $\alpha=0$ 时就得到直线方程(见(2)).这样一来,(4)在 $\alpha\neq 0$ 时表示圆周,在 $\alpha=0$ 时表示直线;直线与圆有了一个统一的复数表示法.

11.1.2　复数的球面表示

取好空间直角坐标系，三个坐标分别为 x_1, x_2, x_3. 考虑半径为 1，中心在原点的球面 S：

$$x_1^2 + x_2^2 + + x_3^2 = 1. \tag{7}$$

点 (0,0,1) 称为北极，记作 N. 复平面 C 等同于 $\{(x_1, x_2, 0): x_1, x_2 \in \mathbf{R}\}$. 复平面 C 交球面 S 于赤道(图 11－1). 现在对 C 中的每一个点 z，将它与北极 N 用直线连接起来，这条直线与球面 S 交于一点 Z. 若 $|z| > 1$，则 Z 在上半球上；若 $|z| < 1$，则点 Z 在下半球面上；当 $|z| = 1$ 时，$Z = z$，它们重合. 这样一来，复平面 C 上的每个点 z 都有球面 S 上的一个点 Z 与它对应. 反过来，球面 S 上的任一点 $Z \neq N$ 与 N 用直线连起来，这连线也必与复平面 C 交于一点 z. 于是除了 $N = (0,0,1)$ 外，复平面 C 与球面 S 上的点都是一一对应的.

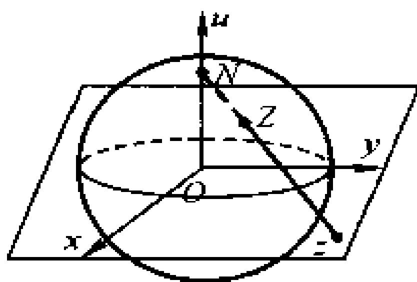


图 11－1

当 $|z|$ 无限增大时，球面上的点 Z 就向 N 无限靠近. 因此，很自然地，在复平面 C 上引进一个理想点作为 N 的对应点，这个点称为无穷远点，记作 $z = \infty$. 加上无穷远点的复平面叫做扩充复平面，记作 $\overline{C} = C \cup \infty$.

$\overline{C}$ 与球面 S 上的点建立起的一一对应称为球极平面投影，S 称为黎曼复球面. 需要强调的是，整个复平面 C 只有一个无穷远点，无穷远点的模是 $+\infty$，辐角是不定的. 在扩充的复平面上，任何一条直线都通过无穷远点.

11.1.3　球极投影的公式

上一段我们讨论了球极投影的几何构造，现在我们来推演这个变换的公式. 也就是要解决下面的问题：已知复数 z，求球面上对应

点 Z 的坐标,以及这个问题的逆.

设复数 z 在复平面 C 上的坐标为 x,y,即 $z=x+\mathrm{i}y$.假定这个复数在球面上的像的坐标是 (x_1,x_2,x_3).我们来求,在空间坐标系中通过 z 和 N 的直线方程.

因为 $(0,0,1),(x,y,0),(x_1,x_2,x_3)$ 这三点在一条直线上,所以它们的坐标满足关系

$$\frac{x_1-0}{x-0}=\frac{x_2-0}{y-0}=\frac{x_3-1}{0-1}.$$

这是两个方程式,从中可以解出 x,y.事实上,由第一比式和第三比式,第二比式和第三比式,分别得出

$$x=\frac{x_1}{1-x_3},y=\frac{x_2}{1-x_3}. \tag{8}$$

从而

$$z=x+\mathrm{i}y=\frac{x_1+\mathrm{i}x_2}{1-x_3}.$$

公式(8)给出了用球面上对应点的坐标来表示平面上点的坐标的公式.为了得到上面公式的逆,我们注意,

$$x^2+y^2=\frac{x_1^2+x_2^2}{(1-x_3)^2}=\frac{1-x_3^2}{(1-x_3)^2}=\frac{1+x_3}{1-x_3}. \tag{9}$$

从而求出

$$x_3=\frac{x^2+y^2-1}{x^2+y^2+1}=\frac{|z|^2-1}{|z|^2+1}. \tag{10}$$

由此可算出

$$1-x_3=\frac{2}{|z|^2+1}.$$

从而由(8)式可得

$$x_1=x(1-x_3)=\frac{2x}{|z|^2+1}=\frac{z+\bar{z}}{|z|^2+1}, \tag{10}$$

$$x_2 = y(1 - x_3) = \frac{2y}{|z|^2 + 1} = \frac{-\mathrm{i}(z - \bar{z})}{|z|^2 + 1}. \tag{10}$$

(10) 给出了球面上的点由对应的复数表示的公式.

11.1.4 球极投影的基本性质

现在我们证明球极投影的一个极重要的性质:

定理 1 在球极投影变换下,复平面上的任何一个圆周都变成球面上的圆周,反之亦然.

注 "圆周"一词要作广义理解,以下把直线视为半径是无穷大的圆周.前面我们已经给出了圆周和直线的统一表示.学了后面的分式线性变换后,我们就会更加体会将圆周和直线统一起来的意义;在分式线性变换下,直线和圆可以互变.

证 在 xy 平面上任一圆周的方程具有形式

$$A(x^2 + y^2) + Bx + Cy + D = 0, \tag{11}$$

其中 A, B, C, D 是实数.当 $A = 0$ 时,(11) 表示一条直线.为了求出球面上的对应曲线,将方程中的 x, y 用 x_1, x_2, x_3 来替换.根据(8)和(9),我们有

$$A\frac{1 + x_3}{1 - x_3} + B\frac{x_1}{1 - x_3} + C\frac{x_2}{1 - x_3} + D = 2,$$

或

$$Bx_1 + Cx_2 + (A - D)x_3 + A + D = 0. \tag{12}$$

方程(12) 是一次的,所以是一张平面.坐标 x_1, x_2, x_3 满足两个方程:(12) 和(7).从而点 x_1, x_2, x_3 在球面(7) 和平面(12) 的交线上,它是球面上的一个圆.

反过来,球面(7) 上的任意一个圆周具有表示

$$Ax_1 + Bx_2 + Cx_3 + D = 0,$$
$$x_1^2 + x_2^2 + x_3^2 = 1.$$

把(10) 代入第一式,得

$$A\frac{2x}{x^2 + y^2 + 1} + B\frac{2y}{x^2 + y^2 + 1} + C\frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} + D = 0.$$

消去分母，合并同类项，得到

$$(C+D)(x^2+y^2)+2Ax+2By+D-C=0.$$

这是复平面上圆的方程. 当 $C+D=0$ 时，方程是一条直线. 证毕.

§11.2 分式线性变换

11.2.1 线性变换

为了给出一个复函数的几何表示，我们考虑两张复平面，一张是 z 平面，一张是 w 平面. 给了一个复变量的复值函数 $w=f(z)$，在几何上就是给出了复平面 z 到复平面 w 的变换或映射：任给一个点 $z\in C$（在 z 平面上），就有一个点 $w\in C$（在 w 平面上）与之对应. 当 z 在 z 平面上变化时，w 在 w 平面上变化. 若 z 在 z 平面上描过一条曲线，则相应地，w 在 w 平面上也描过一条曲线. 但有时我们把两张平面重合在一起，以便立刻看出变换的效果.

我们不研究一般的变换，而研究与我们主题相关的特殊变换.

1) **平移变换** 变换

$$w=z+a$$

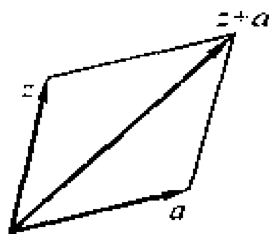


图 11－2

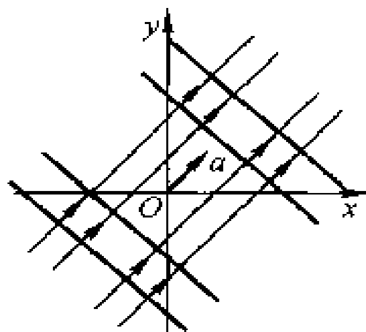


图 11－3

叫做平移变换，这里 a 是常数. 复数加法的几何意义告诉我们，这个

变换把复平面 C 上的任意一点 z,沿着向量 a 所指的方向平移了一个固定距离 $|a|$(图 11 − 2).在平移变换下,平行于向量 a 的直线变为自己,与向量 a 垂直的直线变为另一条与 a 垂直的直线(图 11 − 3).

2) **伸缩变换** 变换

$$w = \rho z, \rho > 0$$

叫做伸缩变换.这个变换把复平面 C 的每一点 z 变到与原点距离为 $\rho|z|$ 的点,并保持辐角不变.当 $\rho > 1$ 时是放大,当 $\rho < 1$ 时是压缩(图 11 − 4).在这种变换下,把以原点为心的圆仍变为以原点为心的圆,把过原点的直线变为自己(图 11 − 5).

3) **旋转变换** 变换

$$w = z e^{i\theta}, \theta \in R$$

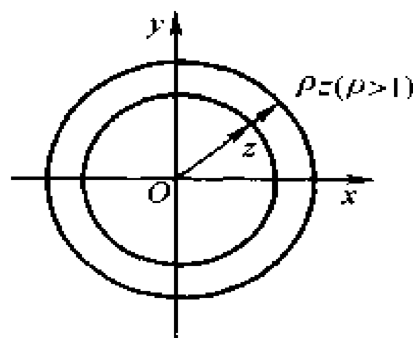


图 11 − 4

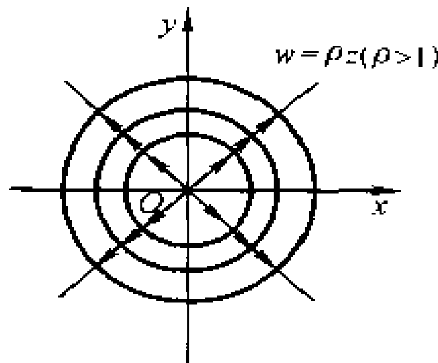


图 11 − 5

叫做旋转变换.在这个变换下,复平面 C 上的每个点绕原点 z 旋转一个角度 θ(图 11 − 6).当 $\theta > 0$ 时,沿反时针方向旋转;当 $\theta < 0$ 时,沿顺时针方向旋转.在旋转变换下,以原点为中心的圆周变为自己;过原点的直线变为另一条过原点的直线(图 11 − 7).

定义 形如

$$w = Az + B \tag{1}$$

的变换称为线性变换.其中 A、B 为复常数.

令 $A = re^{i\theta}, z_1 = e^{i\theta}, z_2 = rz_1$,则

$$w = z_2 + B.$$

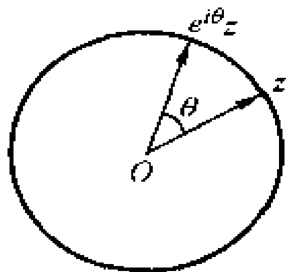


图 11 - 6

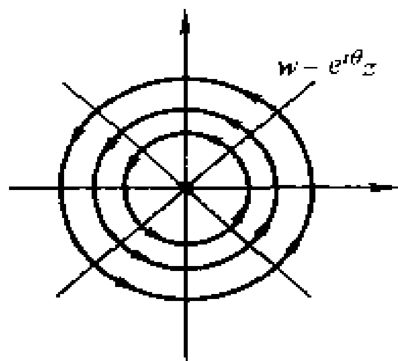


图 11 - 7

由此可见,线性变换是旋转、伸缩和平移三个变换的复合.

对特殊位置的直线和圆,我们证明了线性变换把直线变为直线,把圆变为圆.这个结论对任意位置的直线和圆也对.我们有下面的定理.

定理 1 线性变换(1)把直线变为直线,把圆周变为圆周.

证 直线方程具有形式

$$\overline{\beta}w + \beta\overline{w} + \gamma = 0.$$

将变换(1)代入,得

$$\overline{\beta}(Az + B) + \beta(\overline{Az} + \overline{B}) + \gamma = 0.$$

化为

$$A\overline{\beta}z + \overline{A}\beta\overline{z} + \beta\overline{B} + \overline{\beta}B + \gamma = 0.$$

由于 $A\overline{\beta}$ 与 $\overline{A}\beta$ 互为共轭,而 $\beta\overline{B} + \overline{\beta}B$ 是实数,所以这个方程是直线方程.

圆的证明是类似的,计算稍复杂一些,我们把它留给读者.

11.2.2 反演变换 变换

$$w = \frac{1}{\bar{z}} \tag{2}$$

称为关于单位圆周的反演变换.

令 $w = \rho e^{i\varphi}, z = re^{i\varphi}$ 则

$$\rho e^{i\varphi} = \frac{1}{re^{-i\theta}}.$$

于是

$$\rho = \frac{1}{r}, \varphi = \theta. \tag{3}$$

(3) 的前一式指出,当 $r < 1$ 时,$\rho > 1$;当 $r > 1$ 时,$\rho < 1$. 这就是说,当 z 在单位圆周 $|z| = 1$ 的内部时,w 就在单位圆周 $|z| = 1$ 的外面;当 z 在单位圆周 $|z| = 1$ 的外部时,w 就在单位圆周 $|z| = 1$ 的内部. 又,z 与 w 有相同的辐角,因而它们落在从原点出发的同一射线上(图 11 - 8).

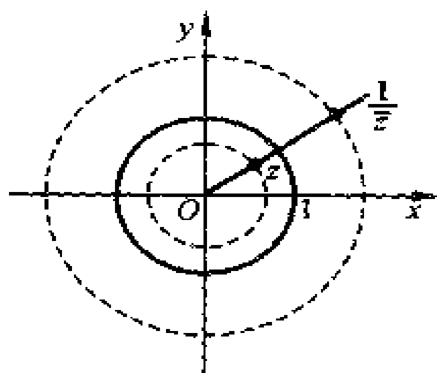


图 11 - 8

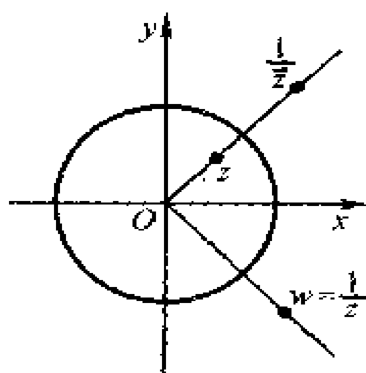


图 11 - 9

这样一来,反演变换(2) 将单位圆内的点变到单位圆外;将单位圆内半径为 r 的圆变为单位圆外半径 $1/r$ 为的圆(设 $r < 1$),同时把单位圆外的圆变到单位圆内,又把过原点的射线变成自己,当 $r = 1$ 时 z

$= w$，即单位圆周变为自己.

11.2.3　倒数变换

变换

$$w = \frac{1}{z} \tag{4}$$

称为倒数变换. 若令 $z' = \overline{z}$，则

$$w = \frac{1}{\overline{z'}}.$$

因而倒数变换是两个变换的复合：1）关于实轴的对称变换（或叫做关于实轴的反射）$z' = \overline{z}$ 和 2）关于单位圆周的反演变换（或称为关于单位圆周的对称或反射）$w = 1/\overline{z'}$. 其几何表示如图（11 − 9）所示.

倒数变换把通过原点的直线仍变为通过原点的直线；把以原点为中心的圆周变为另一个以原点为中心的圆周.

现在问：在倒数变换下，

1）不过原点的直线变成什么？

2）不以原点为心的圆变为什么？

先看直线. 任一直线的方程为

$$\overline{\beta}z + \beta\overline{z} + c = 0, \beta \neq 0, c \in \mathbf{R}.$$

当 $c = 0$ 时，点 $z = 0$ 满足方程，这时直线通过原点. 当 $c \neq 0$ 时，$z = 0$ 不满足方程，这时直线不通过原点. 作倒数变换，即将代入直线方程，得

$$\overline{\beta}\frac{1}{w} + \beta\frac{1}{\overline{w}} + c = 0, c \neq 0 \Rightarrow cw\overline{w} + \beta w + \overline{\beta w} = 0.$$

这是一个圆（条件 § 1 的（5）自然满足，因为这时相当于 γ 的数为 0），而且通过原点. 这就是说，我们有如下结果：

倒数变换将不过原点的直线变为通过原点的圆.

现在看圆. 圆的方程是

$$\alpha z\overline{z} + \overline{\beta}z + \beta\overline{z} + \gamma = 0,$$

其中 α,γ 均为实数，且满足条件

$$\alpha\neq 0,\alpha\gamma<|\beta|^2.$$

今分两种情况进行讨论.

1) $\gamma\neq 0$，这时 $z=0$ 不满足圆的方程，因而圆周不通过原点，把 $w=1/z$ 代入圆的方程，得

$$\alpha\frac{1}{w\overline{w}}+\overline{\beta}\frac{1}{w}+\beta\frac{1}{\overline{w}}+\gamma=0,$$

或

$$\gamma w\overline{w}+\beta w+\overline{\beta w}+\alpha=0.$$

易见，这是一个不过原点的圆(注意，只需将 $\overline{\beta}$ 视为 β，这就是圆的方程，$\alpha\gamma<|\beta|^2$ 的条件也满足). 这样一来，

倒数变换将不过原点的圆变为不过原点的圆.

2) $\gamma=0$，这时 $z=0$ 满足圆的方程. 因此圆周过原点，倒数变换 $w=1/z$ 把原点 O 变到 ∞. 可见在倒数变换下，圆的象曲线不再是圆，它是什么呢?将 $w=1/z$(设 $z\neq 0$)，代入圆的方程，得

$$\alpha\frac{1}{w\overline{w}}+\overline{\beta}\frac{1}{\omega}+\beta\frac{1}{\overline{w}}=0,$$

或

$$\overline{\beta w}+\beta w+\alpha=0$$

这是一条不过原点的直线. 这样一来，

倒数变换将过原点的圆变为不过原点的直线，并且把原点变为 ∞ 点.

前面曾指出，将直线视为半径为 ∞ 的圆，这样一来，我们有

定理 2　变换 $w=1/z$ 将圆周变为圆周. 具体而言

1) 将过原点的直线映为过原点的直线;

2) 将过原点的圆映为不过原点的直线;

3) 将不过原点的直线映为过原点的圆;

4) 将不过原点的圆映为不过原点的圆.

11.2.4 分式线性变换

分式线性变换也称为默比乌斯变换，它是形如

$$w = L(z) = \frac{az + b}{cz + d}, ad - bc \neq 0 \tag{5}$$

的变换.分式线性变换一定可以分解为平移、旋转、伸缩、倒数诸变换的复合.事实上，可分两种情况进行讨论.

1) $c = 0$. 这时(5)取如下形式

$$w = \frac{a}{d}z + \frac{b}{d} = Az + B, A = \frac{a}{d}, B = \frac{b}{d}.$$

这是前面讨论的线性变换，它是平移、旋转及伸缩诸变换的复合.

2) $c \neq 0$. 用多项式除法可得，

$$w = \frac{az + b}{cz + d} = \frac{a}{c} + \frac{bc - ad}{c^2} \cdot \frac{1}{z + \frac{d}{c}}.$$

令

$$z_1 = z + \frac{d}{c}\text{（平移变换）},$$

$$z_2 = \frac{1}{z_1}\text{（倒数变换）},$$

$$w = \frac{a}{c} + \frac{bc - ad}{c^2} z_2\text{（线性变换）}.$$

这时易见，分式线性变换是平移、旋转、伸缩以及倒数变换的复合.

定理 3 任何一个形如(5)的分式线性变换都可分解为平移、旋转、伸缩以及倒数变换的复合.

前面的讨论告诉我们，平移、旋转、伸缩以及倒数诸变换都将直线变为直线，将圆周变为圆周，换言之，将复平面上的圆变为复平面上的圆.这样一来，我们就有下面的重要定理：

定理 4 分式线性变换将复平面上的圆变为复平面的圆.

11.2.5 保角性

分式线性变换的另一个重要性质是它的保角性.

因为线性变换是平移、旋转和伸缩诸变换的复合，而平移、旋转、伸缩诸变换都不改变两直线交角的大小，所以在线性变换下，交角为 θ 的两直线映为交角为 θ 的两直线.

设平面上任意两条曲线交于一点，交点处两曲线都有切线，并形成一个角 θ，这个角 θ 就称为两曲线间在交点处的夹角.

这样一来，线性变换保直线间的夹角也就保持曲线间的夹角.这种保持曲线间夹角的变换称为保角变换，也称为共形映照.线性变换是保角变换.

我们进一步指出，倒数变换 $w = \frac{1}{z}$ 也是保角的.根据前面的讨论，我们只需证明它保持两直线间的夹角就行了，分两种情况讨论.

1）设直线 l 通过原点，它的斜率为 k，变换 $w = 1/z$ 的几何意义指出，l 的象也是一条直线，它关于 x 轴与 l 对称.因此象直线的斜率是 $-k$.这样一来，两条过原点的直线若它们交于角 θ，则它们的像直线也交成角 θ.例如，如图 11 - 10 所示，直线 l_1 与直线 l_2 的夹角为 θ，在倒数变换下，l_1, l_2 的象分别是 l_1' 和 l_2'，l_1' 和 l_2'，的夹角也是 θ.

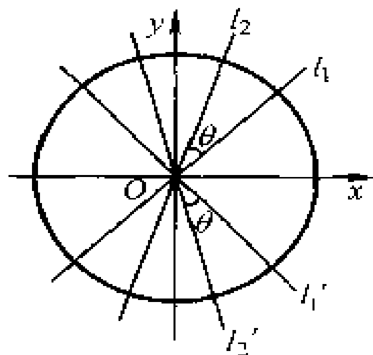


图 11 - 10

2）设直线 l 不通过原点，且具有斜率 k.定理 2 指出，这条直线变为过原点的圆，并且该圆在 $z=0$ 的切线具有斜率 $-k$.所有以 k 为斜率的直线都是彼此平行的，在倒数变换下，它们的象都是在原点彼此相切的圆，除了通过原点的那条直线，它的象直线是那些圆的切线.

这样一来，平面上任意两条交角 θ 的不过原点的直线都被映为交角为 θ 的两个圆.

若两条交角为 θ 的直线，其中一条过原点，另一条不过原点，则

不过原点的一条映为过原点的圆，过原点的一条映为过原点的直线，这条直线与圆的交角仍是 θ.

综上，我们证明了倒数变换是保角的，再结合定理 3，我们就得到如下的定理.

定理 5 分式线性变换是保角变换.

11.2.6 单位圆到自身的分式线性变换

定理 6 若 $|\alpha| < 1, |\beta| < 1$ 则

$$\left|\frac{\alpha-\beta}{1-\overline{\alpha}\beta}\right| < 1. \tag{6}$$

证 证明的思想很简单，只需作些计算.

$$|1-\overline{\alpha}\beta|^2 = (1-\overline{\alpha}\beta)(1-\alpha\overline{\beta}) = 1-\overline{\alpha}\beta-\alpha\overline{\beta}+|\alpha\beta|^2,$$

$$|\alpha-\beta|^2 = (\alpha-\beta)(\overline{\alpha}-\overline{\beta}) = |\alpha|^2-\overline{\alpha}\beta-\alpha\overline{\beta}+|\beta|^2.$$

两者相减，得到

$$\begin{aligned}|1-\overline{\alpha}\beta|^2-|\alpha-\beta|^2 &= 1-|\alpha|^2-|\beta|^2+|\alpha\beta|^2\\ &= (1-|\alpha|^2)(1-|\beta|^2) > 0,\end{aligned}$$

即
$$|\alpha-\beta|^2 < |1-\overline{\alpha}\beta|^2 \Leftrightarrow \left|\frac{\alpha-\beta}{1-\overline{\alpha}\beta}\right| < 1.$$

证毕.

利用定理 6，我们可以证明：

定理 7. 分式线性变换

$$w = e^{i\theta}\frac{z-a}{1-\overline{a}z}, |a| < 1. \tag{7}$$

实现了 $|z| < 1$ 到 $|w| < 1$ 的一一的保角映射.

证 首先，(6) 指出，当 $|z| < 1$ 时，$|w| < 1$，所以变换(7)把 $|z| < 1$ 的点变到 $|w| < 1$ 内.

其次，将(7) 反解出来，得到

$$z = \frac{e^{i\theta}w+a}{1+\overline{a}e^{-i\theta}w}. \tag{8}$$

(8)仍然满足定理6的条件,于是,当$|w|<1$时,$|z|<1$;即逆变换把$|w|<1$内的点映到$|z|<1$内.这就证明了变换(7)是$|z|<1$到$|w|<1$的一一的变换.

最后,保角性是明显的,因为(7)是分式线性变换.

习 题

1. 利用圆的方程完成定理1证明.
2. 利用复合函数微分法求圆

$$A(x^2+y^2)+Bx+Cy+D=0$$

在任意点处的切线的斜率.

§11.3 非欧几何的庞加莱模型

罗巴切夫斯基的几何诞生后长期不为人们所接受,一个重要原因是,它所得到的结论是奇特的,与人们所熟悉的事实大相径庭,尽管在逻辑推理上它是严密的,无懈可击的.但是除去逻辑推理外,人们看不到任何东西.因而在现实空间中找到一个模型来实现它,就变得十分重要了.

第一个这样的模型是1868年意大利数学家贝尔特拉米(Beltrami,1835—1899)给出的.他在罗巴切夫斯基平面的一部分与伪球面的一部分之间建立了点之间的对应,然后用伪球面的内蕴几何来解释罗巴切夫斯基几何.这种解释给予人们很大的启发,使人们对非欧几何有了进一步的认识.这种解释的缺点是,它不是整体的,只是局部的.

第一个整体的罗巴切夫斯基几何的模型是德国数学家克莱因(Felex Klein,1849—1925)给出的.他在1871年的一篇论文中概略地叙述了他的思想.他是第一个认识到无需用曲面来获得非欧几何

模型的人. 克莱因把单位圆作为罗巴切夫斯基的几何的平面，把圆中的弦作为罗巴切夫斯基的几何的直线. 在适当地定义了距离概念之后，使得罗巴切夫斯基的几何的模型得以实现. 克莱因的这一模型使人们对罗巴切夫斯基的几何有了真实感.

克莱因之后，法国数学家庞加莱给出了另一个罗巴切夫斯基的几何的模型. 他把克莱因模型中的弦改为垂直于单位圆周的圆弧，并把非欧几何与分式线性变换联系起来. 庞加莱的模型为罗巴切夫斯基的几何的应用开辟了广阔的道路. 目前这一模型在复分析，黎曼曲面，自守函数，克莱因群等许多数学分支中得到广泛应用.

下面我们就来介绍庞加莱的模型.

11.3.1　非欧平面

用 Δ 表示复平面 C 上的单位圆，即 $\Delta=\{z:|z|<1\}$，它的边界用 $\partial\Delta$ 表示，取 Δ 的内部为非欧平面，任一点 $z\in\Delta$，都是非欧平面内的点，称为非欧点. 在 Δ 内与 $\partial\Delta$ 垂直的圆弧或直线段，称为非欧平面的非欧直线(见图 11 - 11). 由此，所有过原点的直线都是非欧直线.

两条非欧直线间的夹角定义为在交点处它们切线的夹角，若非欧直线是直线段，则切线就是它本身.

我们还需要定义两点间的非欧距离. 任取两个非欧点 $z_1,z_2\in\Delta$，z_1,z_2 间的非欧距离定义为

$$d(z_1,z_2)=\ln\frac{1+\left|\dfrac{z_2-z_1}{1-\bar{z}_1z_2}\right|}{1-\left|\dfrac{z_2-z_1}{1-\bar{z}_1z_2}\right|}\quad(1)$$

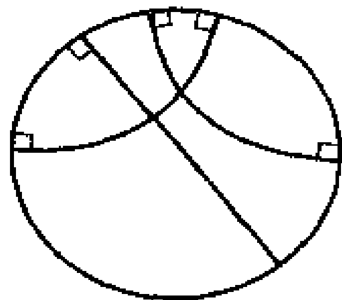


图 11 - 11

这个定义太复杂了！为什么这样定义？分式线性变换就是我们下面要定义的非欧运

动.在欧几里得几何中,刚体运动保持欧氏距离不变.类似地,在非欧几何中,非欧运动保持非欧距离不变.在这一要求下,其距离必然取(1)的形式.我们略去推导过程.

要说明这是距离函数,需要验证它满足通常关于距离的三条性质:

1) $d(z_1,z_2)\geqslant 0$,等号仅在 $z_1=z_2$ 时成立;

2) $d(z_1,z_2)=d(z_2,z_1)$;

3) $d(z_1,z_2)\leqslant d(z_1,z_3)+d(z_3,z_2)$.

我们来分别验证 1) 和 2).3) 的计算复杂故略去.

1) 当 $z_1=z_2$ 时,

$$\left|\frac{z_2-z_1}{1-\overline{z}_1z_2}\right|=0\Rightarrow d(z_1,z_2)=\ln 1=0.$$

当 $z_1\neq z_2$ 时,

$$\left|\frac{z_2-z_1}{1-\overline{z}_1z_2}\right|>0\Rightarrow d(z_1,z_2)>0.$$

2) 只需注意 $|z_1-z_2|=|z_2-z_1|$ 和 $|1-\overline{z}_1z_2|=|1-\overline{z}_2z_1|$ 即可,而这是明显的.

在实轴上任取一点 $r<1$.它和原点间的非欧距离为(图 11－12)

$$d(0,r)=\ln\frac{1+r}{1-r}.\qquad(2)$$

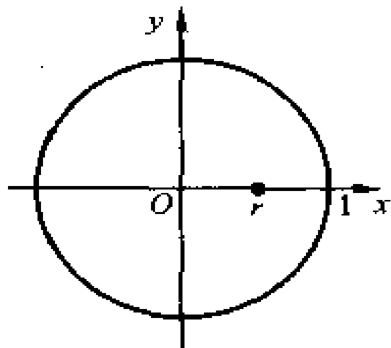


图 11－12

当 $r\to 0$ 时,$d(0,r)\to 0$;当 $r\to 1$ 时,

$$d(0,r)=\ln\frac{1+r}{1-r}\to\infty.$$

这就是说,在非欧平面上,1 是处在无穷远的位置.学了下节讲述的非欧运动之后,读者自己不难证明,$\partial\Delta$ 上的点都处在无穷远的位置.这样我们就在有限的单位圆内表

现了一个无限的非欧空间.

有了非欧距离的概念就能导出非欧圆与非欧圆周的概念.给定 Δ 内一点 z_0 和一个实数 $r>0$,以 z_0 为中心,以 r 为半径的非欧圆定义为集合

$$\Delta = \{z \in \Delta : d(z_0, z) < r\}.$$

到 z_0 的非欧距离等于 r 的一切点的集合

$$\partial\Delta = \{z \in \Delta : d(z_0, z) = r\}.$$

叫做以 z_0 为中心以 r 为半径的非欧圆周.

11.3.2　非欧刚体运动

若 $w = f(z)$ 是 Δ 到 Δ 的一个变换,在这个变换下两点的非欧距离保持不变,即

$$d(z_1, z_2) = d(w_1, w_2),$$

其中 $w_1 = f(z_1), w_2 = f(z_2)$,则称 f 是一个非欧刚体运动.

定理 1　分式线性变换

$$w = \mathrm{e}^{\mathrm{i}\theta}\frac{z-a}{1-\bar{a}z}, |a| < 1. \tag{3}$$

是一个非欧刚体运动.

证　前面已经证明了,(3) 是 Δ 到 Δ 的保角变换.现在只需证明它保持非欧距离.由非欧距离的公式知,只需证明

$$\left|\frac{w_2-w_1}{1-\overline{w_2}w_2}\right| = \left|\frac{z_2-z_1}{1-\bar{z}_1z_2}\right|.$$

事实上,

$$w_1 = \mathrm{e}^{\mathrm{i}\theta}\frac{z_1-a}{1-\bar{a}z_1}, w_2 = \mathrm{e}^{\mathrm{i}\theta}\frac{z_2-a}{1-\bar{a}z_2}.$$

从而

$$w_2 - w_1 = \mathrm{e}^{\mathrm{i}\theta}\frac{z_2-a}{1-\bar{a}z_2} - \mathrm{e}^{\mathrm{i}\theta}\frac{z_1-a}{1-\bar{a}z_1}.$$

提出，通分，化简得

$$w_2 - w_1 = e^{i\theta}(1-|a|^2)\frac{z_2 - z_1}{(1-\bar{a}z_1)(1-\bar{a}z_2)}. \tag{4}$$

又
$$1 - \bar{w}_1 w_2 = 1 - \frac{(\bar{z}_1 - \bar{a})(z_2 - a)}{(1-a\bar{z}_1)(1-\bar{a}z_2)}.$$

通分、化简得到

$$1 - \bar{w}_1 w_2 = \frac{1-|a|^2)(1-\bar{z}_1 z_2)}{(1-a\bar{z}_1)(1-\bar{a}z_2)}. \tag{5}$$

(4) 与(5) 相比，取绝对值，就得出

$$\left|\frac{w_2 - w_1}{1 - \bar{w}_2 w_2}\right| = \left|\frac{z_2 - z_1}{1 - \bar{z}_1 z_2}\right|.$$

证毕．

Δ内任何两点 z_1, z_2，都可经过非欧刚体运动将其中的一点变换为另一点．事实上，变换

$$w = \frac{z - z_1}{1 - \bar{z}_1 z} \tag{6}$$

是一个刚体运动，它把 z_1 变到原点．变换

$$w = \frac{z - z_2}{1 - \bar{z}_2 z}$$

也是一个刚体运动，它把 z_2 变到原点．这个变换的逆变换就将原点变到点 z_2．第一个变换和第二个变换的逆变换复合起来就可把点 z_1 变到点 z_2．

若在变换(6) 下，z_1 的象是 w_1，z_2 的象是 w_2，则

$$w_1 = 0,\ w_2 = \frac{z_2 - z_1}{1 - \bar{z}_1 z_2}.$$

由刚体运动的定义和定理 1 得到

定理 2

$$d(z_1, z_2) = d(0, w_2).$$

经过一个旋转可将 w_2 变到实轴上. 因而任意两点间的非欧距离可化到实轴上来算.

11.3.3　罗巴切夫斯基公理系统

罗巴切夫斯基公理系统指的是欧几里得几何的前四条公设再加上罗巴切夫斯基的第五公设. 我们罗列如下, 并逐条作些说明:

1) 过 Δ 内任意两点可以做一非欧直线段.

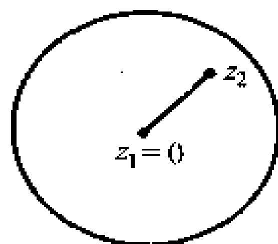


图 11－13

当 z_1, z_2 中有一点是原点时, 连接 z_1, z_2 的非欧线段就是两点间的欧氏线段(图 11－13). 当 z_1, z_2 都不是原点时, 过 z_1, z_2 作一圆弧或直线段, 使之垂直于 $\partial\Delta$. 这是做得到的: 利用分式线性变换(6)就可将 z_1 变到原点, 将 z_2 变到 w_2. 存在连接原点与 w_2 的欧氏线段, (6)的逆变换就将 w_1, w_2 变回到 z_1, z_2; 连接 w_1, w_2 的直线段(延长后就是 Δ 的直径, 垂直于 $\partial\Delta$) 就变成连接 z_1, z_2 的直线段或圆弧. 变换的保角性就保证了连接 z_1, z_2 的直线或圆弧与 $\partial\Delta$ 垂直.

2) 一非欧直线段可以沿两个方向无限延长. 这是明显的, 因为 $\partial\Delta$ 上的点处在无穷远的位置. 事实上, 在分式线性变换下, $\partial\Delta$ 上的点可以互变. 我们已经指出, 1 处在无穷远的位置, 从而 $\partial\Delta$ 上的其它点也处在无穷远的位置.

3) 以任意一点 z_0 为中心, 以任意正数 r 为半径可以作一个非欧圆.

非欧圆具有方程

$$d(z_0, z) \leqslant r.$$

利用距离公式(1)我们有

$$\ln \frac{1+\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|}{1-\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|} \leqslant r \quad \Leftrightarrow \quad \frac{1+\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|}{1-\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|} \leqslant e^r.$$

做简单运算即可解出

$$\left|\frac{z-z_0}{1-\bar{z}_0 z}\right| \leqslant \frac{e^r-1}{e^r+1}.$$

先看 $z_0=0$ 的情况,这时上面的方程化为

$$|z| \leqslant \frac{e^r-1}{e^r+1}.$$

易见,这是以 $z_0=0$ 为中心,以 $\frac{e^r-1}{e^r+1}$ 为半径的欧氏圆.

至于 $z_0 \neq 0$ 的情况,可以借助分式线性变换,将 z_0 点变到原点 O. 求出以 O 为心的圆,再利用这个分式线性变换的逆变换将这个圆变回去,欧氏圆仍变为欧氏圆,不过这时非欧圆的中心不再是欧氏圆的中心. 表达式相当复杂,这里不再赘述.

4) 凡直角都相等.

5) 设 l 是 Δ 内任意一条非欧直线,$z_0 \in \Delta$ 不在 l 上,那么可以过 z_0 作无穷多条非欧直线不与 l 相交.

设 l 与 $\partial\Delta$ 的交点是 A,B,过 z_0 作一条非欧直线 l_1 与 l 相切于点 A,过 z_0 再作另一非欧直线 l_2,与 l 相切于点 B,l_1,l_2 在 Δ 内都不与 l 相交,并且 l_1 与 l_2 所夹的阴影部分任一点与 z_0 所决定的非欧直线都不与 l 相交(图 11 - 14).

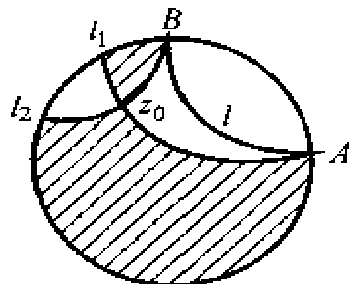


图 11 - 14

l_1,l_2 称为过 z_0 与 l 平行的非欧直线. 其它过 z_0 与 l 不相交的非欧直线叫超平行非欧直线.

这就说明了庞加莱的模型满足罗巴切夫斯基的五条公设.

11.3.4　三角形内角和小于 180°

在欧几里得几何中,利用平行公设可以证明,三角形内角和等于180°.现在非欧几何中平行公设不同了,因而首先想到的是,三角形内角和不会再等于180°.在罗巴切夫斯基几何中,三角形内角和总是小于180°的.

定理 3　非欧三角形的三内角和小于180°.

证　设 w_1, w_2, w_3 是非欧平面内的任意三点,过这三点可以作一个非欧三角形 $\triangle w_1w_2w_3$,如图 11 − 15(a) 所示.Δ到Δ的非欧运动是保角的.因而在非欧运动下,三角形内角和不改变,作非欧运动将非欧三角形 $\triangle w_1w_2w_3$ 变为非欧三角形 $\triangle z_1z_2z_3$,并使 $z_1 = 0$,这时非欧线段 z_1z_2, z_1z_3 在Δ的直径上,然后用欧氏线段连接 z_2, z_3.这一线段与非欧三角形 $\triangle z_1z_2z_3$ 的两直角边构成一个欧氏三角形,设它的内角分别为 α', β', γ'.由图 11 − 15(b) 可看出:$\alpha' = \alpha$, $\beta' > \beta, \gamma' > \gamma$.但 $\alpha' + \beta' + \gamma' = 180°$,所以 $\alpha + \beta + \gamma < 180°$.这就是我们要证的.

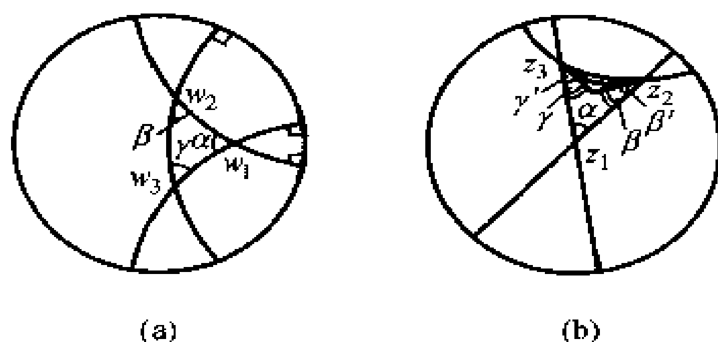


图 11 − 15

11.3.5　真理性讨论

现在在我们的面前出现了两种几何,一种是我们从小就熟悉的

欧氏几何，一种是我们刚刚学过的双曲几何，或罗巴切夫斯基几何．在欧氏几何中平行公理是，过平面上直线外的一点，有且只有一条直线与原直线平行．在双曲几何中，过平面上直线外的一点有无穷多条直线，无论怎样延长也不与原直线相交．这两条公理看来是相悖的．

在我们通常的经验中，如果出现两个相互矛盾的论断，那么一定有一个是错误的．如何判断我们现在遇到的问题呢？是不是一种几何是正确的，而另一种几何是错误的呢？问题没有这样简单．我们遇到了一个远为深刻的问题，必须冲破俗见，寻求对问题的新的理解．答案是，两种几何都是正确的．

那么，判断的标准是什么呢？也就是说，我们根据什么标准去判断一种几何是否正确呢？我们遇到的问题实质上是数学真理与客观真理的关系问题．这是两个在本质上不同的问题．首先，这两种几何是不是数学真理，即在数学上它们是不是正确的．其次，这两种几何是否刻画了我们所生活的物理世界．

第一个问题是逻辑问题，即它们是不是数学真理．这个问题在 19 世纪末已经解决：这两种几何都是数学真理．无论是欧氏几何还是双曲几何在逻辑上都没有错误，都是正确的．那么，为什么这两种几何的结论是如此的不同呢？道理很简单，前提不同结论自然不同．从不同的公理体系出发，自然得出不同的定理．逻辑只看重推论而不看重前提．

当数学家深思新几何出现的意义时，思想得到了一次真正的解放．原来人们可以通过构造不同的公理体系来建立不同的几何学．所以罗巴切夫斯基几何的出现为大量新几何的出现打开了大门．

在罗巴切夫斯基几何诞生的初期，人们总是带着怀疑的态度研究它，企图在这种几何中找出彼此矛盾的命题，以证明罗巴切夫斯基几何不是数学真理．但是没有找到这种命题．人们没有找到这种命题并不表示这种命题就不存在．如果这种研究继续下去，若干年后或许会出现一个智者，他能找到这样的命题，不是仍然说明罗氏几何不是数学真理吗？这说明，人们必须设法证明罗氏几何的公理体系是相容

的,以保证不但现在找不到,以后也找不到这种彼此矛盾的命题.

这是一个生命攸关的问题.数学家必须回答这一问题.否则罗氏几何就始终处在风雨飘摇之中,一阵大风就会把它吹倒.

那么这个问题是由谁解决,又是如何解决的呢?这是在非欧几何发现40年后的事情.那是贝尔特拉米,凯利,F.克莱因和庞加莱等人的工作.办法是在欧氏几何内部建立一个罗氏几何的模型.罗巴切夫斯基几何的庞加莱模型就是其中之一.这就是我们前面所讲的内容.我们的全部推理都是依照欧氏几何进行的.每一个命题都可以解释为非欧几何的命题,也可以解释为欧氏几何的命题.这样一来,如果非欧几何中有一条命题与其它命题矛盾,那么在欧氏几何中也一定有一条命题与其它命题矛盾.于是非欧几何的任何不相容性都会反映出此表示的欧氏几何的对应的不相容性.庞加莱的模型告诉人们,如果欧氏几何是相容的,那么罗氏几何也是相容的.

这种论证法我们称为相对相容性论证.

本来对欧氏几何的公理系统也应该问一问它是不是相容的,即由这个系统出发会不会得到彼此矛盾的命题.但是,欧氏几何的公理系统是如此自然,两千多年以来从未出现过这样的矛盾,所以人们对它深信不疑.而罗氏几何却完全不同,它的命题非常奇特,人们不由得不怀疑它.

罗巴切夫斯基的非欧几何的相容性的一个结果是古老的平行公设的问题的最终解决.相容性确定了这样的事实:平行公设独立于欧氏几何的其它公设.把平行公设作为定理,由其它公设推出它的可能性是不存在的.

非欧几何的相容性还有另一个重要的后果,其影响远远超出了平行公设问题的解决.几何学从传统中解放出来了.接踵而来的是各种各样的几何学的诞生.对整个数学也有类似的影响.数学显现为人类思想的自由创造物.

非欧几何的相容性的研究还大大促进了数学公理系统的研究及几

何基础的研究.希尔伯特和其它一些数学家在这方面作了许多工作.

第二个问题是,我们所居住的物理空间是欧氏的还是非欧的?

这个问题是数学真理与客观真理的关系问题.一切数学真理都存在这一问题.回答这个问题的办法不是靠思辨而是靠实践.但是目前的观测还不能判断我们所生活的空间是欧氏的还是非欧的.

如果观测的办法不能回答这一问题,那么借助物理学的研究成果来探索这一问题将也是一个重要途径.迄今为止,存在两种对物理空间的理解:一种是牛顿的,一种是爱因斯坦的.

对于牛顿,时间和空间形成一个绝对的框架,宇宙的物质活动按照稳定的秩序在其中运行.这个宇宙对每一个观察者都是一样的,不管他站在什么地方或以什么方式旅行.即使对这样的空间,我们也不能判断,它是欧氏的还是非欧的.

1905年,著名物理学家爱因斯坦发表了狭义相对论.按照这种理论,时间与空间是不可分割的.1915年他又发展了广义相对论.在广义相对论中他放弃了对时空均匀性的假定,而认为时空是由不均匀的物质分布和运动所构成.爱因斯坦使光联上时间,时间又联上空间;使能量联上物质,物质联上空间,空间又联上引力.爱因斯坦的物理空间是十分复杂的.所以,从整体上看,这个空间的几何学不会是欧氏的,也不会是罗巴切夫斯基的.

但是在局部上,欧氏几何与非欧几何都是物理空间的很好的近似.

第十二章　微积分前期史

微积分，或者数学分析，是人类思维的伟大成果之一．它处于自然科学与人文科学之间的地位，使它成为高等教育的一种特别有效的工具．遗憾的是，微积分的教学方法有时流于机械，不能体现出这门学科乃是撼人心灵的智力奋斗的结晶；这种奋斗已经历两千五百多年之久，它深深扎根于人类活动的许多领域，并且，只要人们认识自己和认识自然的努力一日不止，这种奋斗就将继续不已．

R．柯朗

课本中的字斟句酌的叙述，未能表现出创造过程中的斗争、挫折，以及在建立一个可观的结构之前，数学家所经历的艰苦漫长的道路．学生一旦认识到这一点，他将不仅获得真知灼见，还将获得顽强地追究他所攻问题的勇气，并且不会因为他自己的工作并非完美无缺而感到颓丧．实在说，叙述数学家如何跌交，如何在迷雾中摸索前进，并且如何零零碎碎地得到他们的成果，应能使搞研究工作的任一新手鼓起勇气．

M．克莱因

温故而知新，可以为师矣．

论语

学习微积分概念的发展史将会使我们受益良多．

17世纪正是由中世纪向新时代过渡的时期．资本主义开始发展，并成为与封建制度作斗争的先进力量．精密科学从当时的生产与社会生活中获得巨大动力．航海学引起了对天文学及光学的高度兴趣．造船学，机器制造与建筑，堤坝及运河的修建，弹道学及一般的军事问题等等，促进了力学的发展．天文学，力学，光学以及工业技术本

身,又要求对当时的数学作彻底的革新.

革新的旗帜是变量,有了变量数学才能研究运动与变化,才能适应新时期科学技术对数学的新要求.科学技术及自然科学方面提出的新问题导致了无穷小量的研究,从而诞生了数学分析这一学科.

研究的开始是"手工业"式的.建立每一个个别的结果都要采取特殊的办法.随着时间的推移情况逐渐有所改变,终于出现了用一般的方法去解同一类型的问题,建立了各类问题之间的联系,弄清楚了一些基本概念.而且微分学与积分学是相互独立地发展起来的,最后在牛顿和莱布尼茨手中建立了微分学和积分学的联系,完成了微积分的创立.

那么,促使微积分产生的主要因素是什么呢?当时科学面临的主要问题是什么呢?微积分的创立首先是为了处理下列四类问题.

1.已知物体运动的路程与时间的关系,求物体在任意时刻的速度和加速度.反过来,已知物体运动的加速度与速度,求物体在任意时刻的速度与路程.

困难在于,17世纪所涉及的速度和加速度每时每刻都在变化.计算平均速度可用运动的时间去除运动的距离.但对瞬时速度,运动的距离和时间都是0,这就碰到了0/0的问题.这是人类第一次碰到这样的问题.

2.求曲线的切线.这是一个纯几何的问题,但对于科学应用具有重大意义.例如在光学中,透镜的设计就用到曲线的切线和法线的知识.在运动中也遇到曲线的切线问题.运动物体在它的轨迹上任一点处的运动方向,是轨迹的切线方向.

实际上,"切线"本身的意义也是没有解决的问题.对于圆锥曲线,把切线定义为和曲线只接触一点而且位于曲线一边的直线就足够了;这个定义古希腊人已经知道.但是对于17世纪所用的比较复杂的曲线,它就不适用了.

3.求函数的最大值和最小值问题.在弹道学中这涉及到炮弹的

射程问题. 在天文学中涉及到行星和太阳的最近和最远距离问题.

4. 求积问题. 求曲线的弧长,曲线所围区域的面积,曲面所围的体积,物体的重心等. 这些问题在古希腊已开始研究,但他们的方法缺乏一般性.

§ 12.1　积分学的早期史

12.1.1　欧多克索斯的穷竭法

我们由积分学的早期史开始. 讲到面积、体积和弧长的计算,图形重心的定位,这实际上要追溯到遥远的古希腊.

苏格拉底的同时代人,巧辩家安提丰(约公元前 500 年)是对圆的求积问题作出贡献的第一人. 安提丰提出,随着一个圆的内接正多边形的边数逐次成倍增加,圆与多边形面积的差将被穷竭. 安提丰的论断包含了希腊穷竭法的萌芽. 但穷竭法通常以欧多克索斯命名. 欧多克索斯(Eudoxus 公元前 400 – 公元前 350) 是古希腊柏拉图时代最伟大的数学家和天文学家. 生于小亚西亚西南的克尼图斯. 他假定量是无限可分的,并以下述命题为基础.

命题 1　如果从任一量中减去不小于它的一半的部分,从余量中再减去不小于它的一半的另一部分,如此继续下去,则最后留下一个小于任何给定的同类量的量.

为了更好地理解穷竭法,我们来详细地考察一个例子.

命题 2　圆的内接相似正多边形面积之比等于圆的直径的平方之比.

这个证明是大家熟知的,故略去. 看下面的命题.

命题 3　圆与圆的面积之比等于其直径平方之比.

证　设 A_1, A_2 分别是直径为 d_1, d_2 的两个圆的面积. 我们要证

$$A_1 : A_2 = d_1^2 : d_2^2.$$

在图 12 – 1 中，设 AB 为圆的内接正多边形的一边，M 为$\overset{\frown}{AB}$的中点．RS 在点M 与圆相切．因为 $\triangle AMB$ 的面积是矩形 $ABSR$ 的面积的一半，所以它也大于弓形 AMB 面积的一半．这样，当把内接正多边形的边数不断地加倍时，我们就能使多边形与圆的面积之差小于任意小的给定面积．

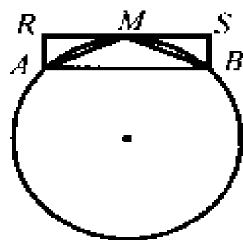


图 12 – 1

回到我们的命题．假定命题不成立．先设

$$A_1 : A_2 > d_1^2 : d_2^2.$$

我们可以在第一个圆中作一个内接正多边形，使其面积 P_1 与 A_1 之差如此之小，使得

$$P_1 : A_2 > d_1^2 : d_2^2.$$

在第二个圆中作一个内接正多边形，与第一个内接正多边形相似，使其面积为 P_2．根据命题 2，

$$P_1 : P_2 = d_1^2 : d_2^2.$$

因此，

$$P_1 : A_2 > P_1 : P_2 \quad \Rightarrow \quad P_2 > A_2.$$

这就得到一个矛盾；因为正多边形的面积不超过它的外接圆的面积．

类似地，我们可以证明，不可能有

$$A_1 : A_2 < d_1^2 : d_2^2.$$

于是命题得证．这里使用的方法叫双归谬法．

双归谬法是严谨的，但不能得出成果．换句话说，一旦知道了一个公式，双归谬法就能提供证明它的灵巧工具，然而这方法对结果的最初发现不起作用．在这方面它很像归纳法．

欧多克索斯还证明了棱锥体积是同底同高的棱柱体积的三分之一，以及圆锥体积是同底同高的圆柱体积的三分之一．但他没有明确的极限思想．

12.1.2 阿基米德的平衡法

在古人中，阿基米德对穷竭法作出了最巧妙的应用.阿基米德大约于公元前287年出生在西西里岛的叙拉古.叙拉古是当时希腊的一个殖民城市.公元前212年罗马人攻陷叙拉古时阿基米德被害.城被攻破时，他正在潜心研究画在沙盘上的一个图形.一个刚攻进城的罗马士兵向他跑来，身影落在沙盘里的图形上，他挥手让士兵离开，以免弄乱了他的图形，结果那士兵就用长矛把他刺死了.

阿基米德的死象征一个时代的结束，代之而起的是罗马文明.

阿基米德有十部著作流传至今，有迹象表明他的另一些著作失传了.现存的这些著作都是杰作，计算技巧高超，证明严格，并表现了高度的创造性.在这些著作中，他对数学作出的最引人注目的贡献是，积分方法的早期发展.

在阿基米德《论球和柱体》一书中，第一次出现了球和球冠的表面积，球和球缺的体积的正确公式.《论球和柱体》一书分为两卷.在第一卷的命题33和34的推理中，他指出，如果圆柱的底等于球的大圆，圆柱的高等于球的直径，则球的表面积恰好等于圆柱的总面积(包括侧面积和两底的面积)的2/3，圆柱的体积恰好等于球的体积的3/2.由此不难得出我们熟知的公式：

$$S = 4\pi r^2, V = \frac{4}{3}\pi r^3,$$

其中 S 和 V 分别表示半径为 r 的球的表面积和体积.

这些结果是通过一系列命题一步一步推导出来的，这个过程蕴含着积分思想.

阿基米德的另一短论“方法”是1906年才发现的.这个短论在形式上是致亚历山大里亚大学依拉托斯芬书的一封信.在这个短论中，阿基米德说，他以特殊的方法得出了他的结果，其中形式上利用了杠杆平衡理论，但本质上是含有由线组成平面图形，由平面组成立体的思想.这种借助“原子论”方法找到的真理，阿基米德用反证法给出

了严格的证明.

为了具体说明这种方法,我们来应用这种方法求球的体积.圆柱的体积和圆锥的体积比较好求,这在阿基米德时代早已知道.求球的体积要困难得多.阿基米德借助圆柱和圆锥的体积求出了球的体积.

命题 4　半径为 r 的球的体积是 $V = 4\pi r^3/3$.

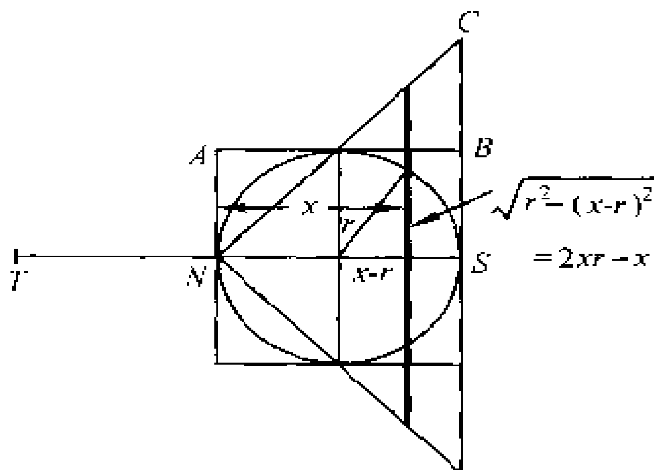


图 12－2

解　把球的直径放在 x 轴上.设 N 是它的北极,S 是它的南极,且原点与北极重合(图 12－2).画出 $2r \times r$ 的矩形 $NSBA$ 和 $\triangle NSC$.绕 x 轴旋转矩形 $NSBA$ 和 $\triangle NSC$,得到一个圆柱体和一个圆锥体.圆的旋转得到球体.然后从这三个立体上切下与 N 的距离为 x,厚为 Δx 的竖立的薄片.

这些薄片的体积近似为

球体:$\pi(2xr - x^2)\Delta x$;

柱体:$\pi r^2 \Delta x$;

锥体:$\pi x^2 \Delta x$,

取出球体和锥体的薄片,把它们的质心吊在点 T,

使 $TN = 2r$.这两个薄片绕 N 的合成力矩为

$$[\pi(2xr - x^2)\Delta x + \pi x^2\Delta x]2r = 4\pi r^2 x\Delta x.$$

不难看出,这是圆柱割出的薄片处于原来位置时绕 N 的矩的四倍.把所有这样割出的薄片绕 N 的力矩加在一起,我们便得到

$$2r[\text{球的体积} + \text{圆锥的体积}] = 4r[\text{圆柱的体积}]$$

即

$$2r\left[\text{球的体积} + \frac{8\pi r^3}{3}\right] = 8\pi r^4.$$

由此我们就求出了球的体积

$$\text{球的体积} = \frac{4\pi r^3}{3}.$$

这就是阿基米德求球的体积的方法.

阿基米德的数学素养极高,他决不把这种方法当作证明,而是随后利用穷竭法给出了一个严格的证明.数学史家克莱因曾这样评论道,阿基米德的严格性比牛顿和莱布尼茨的要高明得多.

在阿基米德的平衡法中,他把一个量看成由大量的微元所组成,这与现代的积分法实质上是相同的.阿基米德的著作是希腊数学的顶峰.

阿基米德对他在《论球和圆柱》一书中作出的贡献十分满意,以致于他希望在他死后把一个内切于圆柱的球的图形(图 12-3)刻在他的墓碑上.后来当罗马将军马塞拉斯得知阿基米德在叙拉古陷落期间被杀的消息时,他为阿基米德举行了隆重的葬礼,并为阿基米德立了一块墓碑,上面刻着阿基米德生前要求的那个图形,以此来表示他对阿基米德的尊敬.这块墓地后来湮没了.令人惊奇的是,在 1965 年,当为一家新建的饭店挖地基时,铲土机碰到一块墓碑,上面刻着一个内切于圆柱的球的图形.叙拉古人又为他们的这位伟人重建了茔墓.

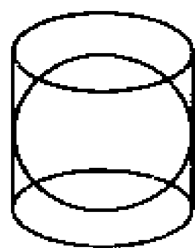


图 12-3

12.1.3　不可分素方法

第一个试图阐明阿基米德方法，并将他的方法给予推广的是德国的天文学家和数学家刻卜勒. 刻卜勒在 1615 年写了一本书名为《酒桶的新立体几何》，书中包含用无穷小元素求面积和求体积的许多问题，其中有 87 种新的旋转体的体积. 刻卜勒的工作的直接继承者是 B. 卡瓦列里(B. Cavalieri).

卡瓦列里于 1598 年生于意大利的米兰. 他是伽利略的学生. 从 1629 年起他一直担任波洛尼亚的大学教授，于 1647 年谢世，只活了 49 岁. 他对数学的最大贡献是 1635 年发表的关于不可分素法的专论，名为《不可分素几何学》(Geometria indivisibilibus).

卡瓦列里说："要决定平面图形的大小可以用一系列平行线；我们设想在这些图形上画了无穷多平行线"(图 12 - 4). 他以同样的方式处理了立体，只是那里不是直线，而是平面. 这些直线(或平面)就是不可分素. 他的不可分素法写得晦涩难懂，使人难以确切理解"不可分素"倒底是什么.

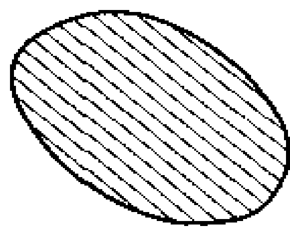


图 12 - 4

卡瓦列里利用不可分素法解决了整数幂的幂函数的积分问题. 用现代的语言说，他算出了下面的积分：

$$\int_0^a x^m dx = \frac{1}{m+1} a^{m+1}. \tag{1}$$

卡瓦列里比刻卜勒进了一步：刻卜勒每次只能算具体的体积，而没有形成一个一般的方法.

把卡瓦列里的结论稍加整理就得出卡瓦列里原理：

1. 如果两个平面片处于两条平行线之间，并且平行于这两条平行线的任何直线与这两个平面片相交，所截二线段长度相等，则这两个平面片的面积相等.

2. 如果两个立体处于两个平行平面之间，并且平行于这两个平

面的任何平面与这两个立体相交,所得二截面面积相等,则这两个立体的体积相等.

卡瓦列里原理是计算面积和体积的有用工具.它的基础很容易用现代的微积分严格化.承认这两个原理我们就能解决许多求积问题.今举两个例子来说明卡瓦列里原理的应用.

例 求椭圆的面积.

解 在直角坐标系中,圆和椭圆分别有方程

$$x^2+y^2=a^2;\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,a>b.$$

图 12－5 画出了它们的图形.由每个方程解出 y,我们分别得到

$$y=\sqrt{a^2-x^2},$$

$$y=\frac{b}{a}\sqrt{a^2-x^2}.$$

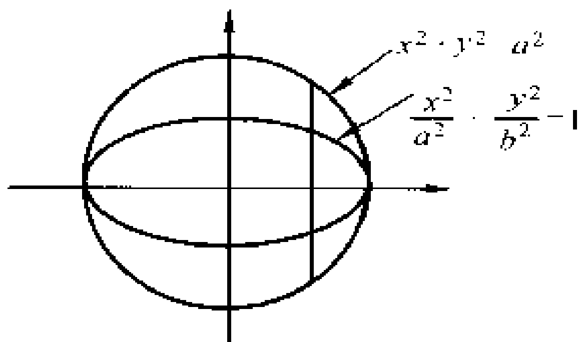


图 12－5

由此可知,椭圆和圆的纵坐标之比是 b/a.所以,椭圆和圆的相应弦之比也是 b/a(这里对原理作了广义的理解:从相等推广为成比例).因此,根据卡瓦列里原理1,椭圆和圆的面积之比也是 b/a.我们得到结论:

$$椭圆面积=b/a(圆面积)=(b/a)(a^2\pi)=ab\pi.$$

刻卜勒也是用这种方法求的椭圆面积.

例 求半径为 r 的球的体积.

解 在图 12－6 中,左边是一个半径为 r 的半球,右边是一个半径为 r 高为 r 的圆柱和一个以圆柱的上底为底、以圆柱的下底中心为顶点的圆锥.这个半球和挖出圆锥的圆柱处在同一平面上.这时用平行于底面、与底面距离为 h 的平面截两个立体.所得截面一个是圆形,一个是环形.用初等几何不难证明,这两个截面的面积都等于

$\pi(r^2-h^2)$. 根据卡瓦列里原理 2,这两个立体有相等的体积.所以球的体积为

$$V=2(\text{圆柱的体积}-\text{圆锥的体积})=2\left(\pi r^3-\frac{\pi r^3}{3}\right)=\frac{4}{3}\pi r^3.$$

利用卡瓦列里原理可以简化中学立体几何课程中许多体积公式的推导过程.这种方法在西方已为许多作者接受,并且在教学法的立场上受到人们的支持.

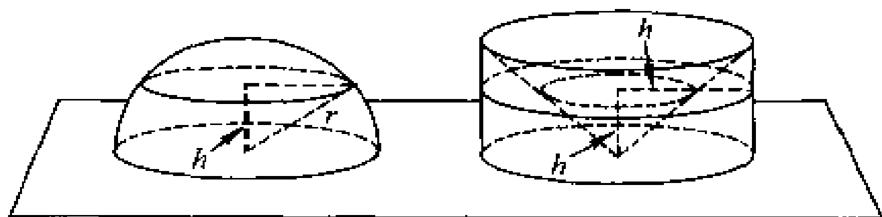


图 12－6

12.1.4 不可分素方法的进一步发展

卡瓦列里的不可分素方法引起了很大的争论,也得到很大的发展.从法国数学家费马的通信中可看出,他也得到了卡瓦列里的一般结果,比卡瓦列里还要早一些.还应该提到英国数学家沃利斯(John Wallis 1616—1703)和他的著作《无穷数量的算术》.他把计算联系到自然数的方幂和的问题.在他的著作中明白地提出了极限过程.

更接近于定积分的现代理解法的是法国数学家、物理学家和哲学家帕斯卡(Blaise Pascal, 1623—1662).他计算了种种面积、体积、弧长,并解决了求重心位置等一系列问题.

12.1.5 刘徽的贡献

中国古代数学家对微积分的贡献很少为世人所知,但是中国古代数学家对微积分的确作出了重大贡献,我们需要在这里花一点笔墨介绍一下.

刘徽是中国数学史上非常伟大的数学家，活动于魏晋间. 他是中国古典数学理论的奠基者之一. 他的杰作《九章算术注》和《海岛算经》现在有传本，是我国最可宝贵的数学遗产. 遗憾的是，关于这位伟大数学家的籍贯、履历和生卒年代，我们没有可靠的史料，无法写出一篇简要的传记.

刘徽全面论述了《九章算术》所载的方法和公式，指出并纠正了其中的错误. 刘徽对积分学的贡献主要有两点.

1) 刘徽的割圆术是他的最著名的一项工作. 他应用极限思想证明了求圆面积公式和给出了计算圆周率的方法. 他从圆内接正 6 边形开始，依次得正 12 边形，正 24 边形 ……. 割得越细，正多边形的面积与圆的面积之差就越小. 他说，"割之弥细，所失弥少. 割之又割以至于不可割，则与圆周合体而无所失矣". 他得到的圆周率是 $3927/1250(\approx 3.1416)$. 他提出的计算圆周率的科学方法奠定了此后千余年中国圆周率计算在世界上的领先地位.

2) 关于解决体积问题的设想.《九章算术》中已有了求球体积的公式，相当于 $V=\frac{9}{16}d^3$（这里 d 是直径）. 刘徽指出这个公式是错误的. 其原因在于错误地把球与外切圆柱的体积的比看成 $3:4$. 为了推导体积公式，刘徽在正方体内作了两个相互垂直的圆柱，并称两圆柱的公共部分为"牟合方盖"（图 12－7）. 他虽未能完成球的体积的推导，但他正确地指出，"牟合方盖"与其内切球体体积之比为 $4:\pi$. 在算法理论和数学思想方面都给后人以极大的启发. 他写了如下一段文字描述他曾做的努力：

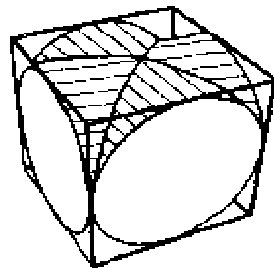


图 12－7

观立方之内，合盖之外，虽衰杀有渐，而多少不掩. 判合总结，方圆相缠，浓纤诡互，不可等正. 欲陋形措意，惧失正理. 敢不阙疑，以俟能言者.（参考译文：考查立方体之内，牟合方盖之外，虽然有部分体

积已被消去，但仍不足将圆锥比球多出的体积完全去掉．剖分与合并的最终结果，方与圆相纠结，超出与不足相混杂，不能得到一个规范的形状．想要凭借我的浅薄学识给出一个解答，又怕背离了正确的原理．敢不存疑，以待能者，作出正确的解答．）

12.1.6 祖暅原理

二百年后，祖冲之的儿子祖暅沿着刘徽的思路完成了球体公式的推导．

祖暅字景烁，是南北朝时南朝著名的数学家和天文学家．在梁朝作过员外散骑侍郎、太府卿、南康太守等官职．他从小就受到良好的家庭教育，青年时代已经对天文学和数学有了很深的造诣，是祖冲之科学事业的继承者．《缀术》就是他们父子共同完成的数学杰作．

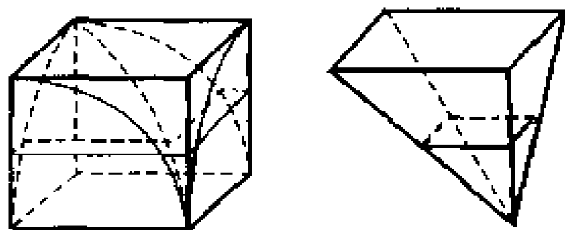


图 12－8

左图虚线部分为牟合方盖的 1/8，因左图阴影部分与右图阴影部分的面积相等．据祖暅原理，故其体积等于正方形与四棱锥的体积之差

在推导“牟合方盖”体积的过程中，祖暅提出了“幂势既同，则积不容异”的原理，后来称为“祖暅原理”．用现代语言来说，就是“若两立体在等高处具有相同的截面面积，则这两立体的体积相等”．这就是前面提到的卡瓦列里原理，但比卡瓦列里早了一千年．根据祖暅原理，可将“牟合方盖”的体积化成一个正方体和一个四棱锥的体积之差（图 12－8）．由此求出“牟合方盖”的体积等于 $\frac{2}{3}d^3$ 并得到球的体

积为$\frac{1}{6}\pi d^3$，这里 d 是球的直径.

顺便指出，祖暅还发现了北极星与北天极相差一度有余，纠正了北极星就是天球北极的错误观点.

上面我们概括地介绍了积分学的早期发展史.这段历史纵跨了二千年的时间.相对来说，微分学的历史就短的多.为什么呢？原因是积分学研究的问题是静态的，而微分学则是动态的，它涉及到运动.在生产力还没有发展到一定阶段的时候，微分学是不会产生的.

§12.2　微分学的早期史

在 17 世纪，由于两位杰出的数学家伽利略和刻卜勒的一系列发现，导致了数学从古典数学向现代数学的转折.在 25 岁以前伽利略就开始作了一系列实验，发现了许多有关物体在地球引力场运动的基本事实.刻卜勒在 1619 年前后归纳出著名的行星运动三定律.这些成就对后来的绝大部分的数学分支都产生了巨大影响.伽利略的发现导致了现代动力学的诞生，刻卜勒的发现则产生了现代天体力学.这些学科的发展都需要一种新的数学工具，这就是研究运动与变化过程的微积分.

有趣的是，积分学的起源可追溯至古希腊时代，但直到 17 世纪微分学才出现重大突破.微分学主要来源于两个问题的研究，一个是作曲线切线的问题，一个是求函数的最大、最小值的问题.这两个问题在古希腊也曾考虑过.例如，在古希腊就能作出圆和圆锥曲线的切线.阿波罗尼奥斯在他的《圆锥曲线》一书中讨论过圆锥曲线的法线，把它当作从一点至曲线的最大和最小线段.在古希腊的著作中也可以找到对极大、极小问题的讨论.但古希腊对这两个问题的讨论远不及对面积、体积、弧长问题讨论得那么广泛和深入.

12.2.1 费马以前的工作

从一般意义上重新讨论曲线的切线问题由法国数学家罗贝瓦尔(G.P. de Roberval 1602—1675)提出.他认为,曲线是由运动的点生成的,点的运动又可以分解为两个已知的运动.两个已知的运动的速度向量的合成向量给出曲线的切线.一个例子是求抛物线的切线.

例 求抛物线的切线.

解 抛物线上的点可视为由两个运动组成的.一个是离开焦点的运动,一个是离开准线的运动.因为运动的点到准线和焦点的距离始终是相等的,所以两个运动的速度向量也相等.因此,抛物线上任一点处的切线是由焦点至该点的射线和通过该点的准线的垂线所组成的角的平分线,如图 12－9 所示.

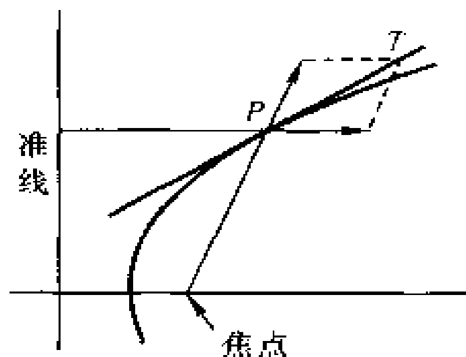


图 12－9

意大利物理学家和数学家托里拆利(Torricelli 1608—1647)也持有这种观点.后来出现了发明权的争论.这个方法是有价值的,因为它把纯几何与物理联系起来了,这在过去是没有的.但是这个定义在数学上不能令人满意,因为它用物理概念定义切线,而许多曲线与物理无关,这个定义就不适用了.所以,看起来很漂亮但不实用,换一种曲线就不行了.

另一个作切线的方法是笛卡儿给出的.写在他的《几何学》一书的第二部分.他的方法仅限于代数曲线,并且会遇到代数上的困难.

12.2.2 费马求极大、极小值的方法

上面的方法有一定的局限性,不能用到一般的情形,也没有包含可能产生微分学的方法.属于微分方法的第一个真正值得注意的先驱工作是 1629 年费马给出的.不过这些思想直至八、九年后才较多

地为人所知.刻卜勒已经观察到,一个函数的增量通常在函数的极大值或极小值处变得无限地小.费马利用这一事实找到了求极大值和极小值的方法.他的方法如下:

设 $f(x)$ 在 x 处有极大值或极小值,并设 e 是一个很小的量,那么 $f(x+e)$ 的值几乎等于 $f(x)$ 的值.因此我们可以先假定它们相等 $f(x+e)=f(x)$,然后让 e 等于 0,等式仍相等.消去 e,得一方程,这个方程的根就是使 $f(x)$ 取极大值或极小值的 x.我们举一个例子来说明这种方法.

例 将一个常量 M 分成两部分,使其乘积最大.

解 费马用大写字母表示变量.设一部分为 A,则另一部分就是 $M-A$.这样,所求的乘积用函数 $f(A)=A(M-A)$ 表示.问题化为 A 取什么值 $f(A)$ 最大.依费马的方法,作乘积

$$(A+e)[M-(A+e)],$$

使它等于 $A(M-A)$,我们有

$$A(M-A)=(A+e)[M-(A+e)].$$

化简得到

$$Me-2Ae-e^2=0.$$

消去 e,得

$$M-2A-e=0.$$

再让 $e=0$,我们得到 $2A=M$.所以答案是,两部分都是 M 的一半.

费马解释的逻辑有待改进,但他的方法等价于:假定

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=0.$$

即假定 $f(x)$ 的导数等于 0.这就是我们在微积分中学过的方法.

费马的方法除了逻辑上的不完整外,还有另外两个问题:一个是费马的方法对极大值与极小值未加区别;一个是他不知道 $f(x)$ 的导数为 0 只是极值的必要条件而非充分条件.这也难怪,因为他的工作只是个起点.

12.2.3　费马求切线的方法

费马还创造了求曲线切线的方法.

设一曲线在直角坐标系中的方程为 $f(x,y)=0$. 现在要求这条曲线在某一点处的切线. 费马的办法是先求该点的次切线. 次切线指的是 x 轴上两点间的一个线段. 其中一点是从该点向 x 轴作垂线所得到的垂足 A, 另一点是切线与 x 轴的交点 P(图 12－10). 设 $PA=t$. 如果能求出次切线 t 的长度, 那么点 P 的位置就完全确定了, 从而切线 PC 就可以画出来了.

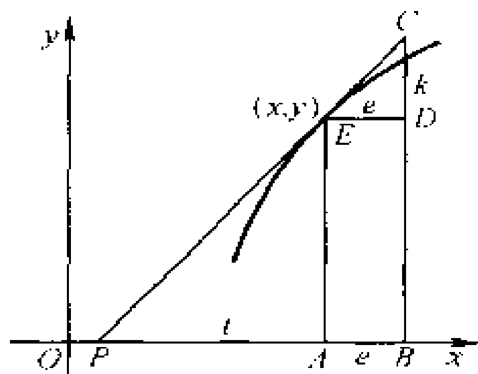


图 12－10

如何确定次切线 t 呢? 费马采用了与求函数的极大、极小值类似的方法. 我们的目的是求曲线在点 E: (x,y) 的切线. 费马考虑切线上邻近的一点 C. 点 C 的 x 坐标是 $x+e$. 点 C 的 y 坐标可用相似三角形的定理求出. 事实上, $\triangle CED \sim \triangle EPA$, 从而有

$$\frac{k}{e}=\frac{y}{t} \quad \Rightarrow k=\frac{ye}{t} \quad \Rightarrow CB=k+y=y\left(1+\frac{e}{t}\right),$$

这样, 在切线上邻点 C 的坐标为 $\left(x+e, y\left(1+\frac{e}{t}\right)\right)$. 费马暂时认为这一点也在曲线上, 于是有

$$f\left(x+e, y\left(1+\frac{e}{t}\right)\right)=0$$

(当 $e=0$ 时这个等式自然是成立的). 然后解此方程, 令 $e=0$, 就可求出 t. 下面用例子说明这种方法.

例　求笛卡儿叶形线 $x^3+y^3=nxy$ 在一点 (x,y) 处的次切线.

解　仿上,我们有

$$(x+e)^3+y^3\left(1+\frac{e}{t}\right)^3=ny(x+e)\left(1+\frac{e}{t}\right).$$

展开,合并同类项,得到方程

$$e\left(3x^2+\frac{3y^3}{t}-\frac{nxy}{t}-ny\right)$$

$$+e^2\left(3x+\frac{3y^3}{t^2}-\frac{ny}{t}\right)+e^3\left(1+\frac{y^3}{t^3}\right)=0.$$

现在用 e 除上式,然后令 $e=0$,就得到

$$\left(3x^2+\frac{3y^3}{t}-\frac{nxy}{t}-ny\right)=0,$$

即

$$t=-\frac{3y^3-nxy}{3x^2-ny}.$$

这样我们就求出了点 (x,y) 处的次切线.有了次切线,切线也就不难求出了.

曲线的切线问题和函数的极大、极小值问题都是微分学的基本问题.正是两个问题的研究促进了微分学的诞生.费马在这两个问题上都作出了重要贡献.从上面的分析可看出,费马处理这两个问题的方法是一致的.用现代语言来说,都是先取增量,而后让增量趋向于0.而这正是微分学的实质所在,也正是这种方法不同于古典方法的实质所在.

12.2.4　费马在积分学方面的贡献

与 §12.1 表达式(1) 等价的命题除了卡瓦列里之外,还见于托里拆利、罗贝瓦尔和帕斯卡的著作中.费马也给出了这个法则的几种证明.和其它人比起来,他的工作是领先的.在 1644 年前,他已发现了关于分数幂的"抛物线"$a^my^n=b^nx^m$ 的求面积、体积、及其重心的方法.

在费马求面积的过程中,我们看到了定积分概念与运算的大部

分的主要方面.他把曲线下的面积分割为小的面积元素,利用矩形和曲线的解析方程,求出这些和的近似值,以及在元素个数无限增加,而每个元素面积成为无限小时,将表达式表示为和式极限的方式.可以说,费马认识到除了积分的抽象定义本身以外的所有方面.但是,他没有认识到所进行的运算本身的重要意义.对他来说,这个运算正如对他的前人一样,只是求面积的问题,只是回答一个具体的几何问题.只有牛顿和莱布尼茨才把这一问题上升到一般概念,认为这是一种不依赖于任何几何的或物理的结构性运算,并给予特别的名称.

费马还考虑了求抛物体的重心问题.他得到的结果当然是早就知道的.在一千九百多年以前,阿基米德在他的《方法篇》中已算出这一结果.而且在一个世纪以前又为康曼第努和麦洛里克斯重新发现过.费马的贡献在于,他第一次采用了相当于今天的微分学中的方法,而不是类似于积分求和的方法.一个通常用求和的方法得到的结果,竟能用求极大、极小值的方法得到,这使他的朋友罗贝瓦尔感到惊奇.奇怪的是,他用求极大、极小值的方法求重心,竟然没有看到这两类问题——微分学问题与积分学问题——的基本联系.只要费马能对他的抛物线和双曲线求切线和求面积的结果更仔细地考察一下,他就可能发现微积分基本定理.

费马当然在某种意义下理解到这两类问题有一个互逆关系.他之所以没有作进一步的考虑,可能是由于他以为他的工作只是求几何问题的解,而不是本身就很有意义的推理过程.他的极大、极小值方法,切线方法及求面积的方法,在他看来是解决这些问题的特有的方法,而不是新的分析学.此外,在应用上也有局限性.费马只知道把它们应用到有理式的情况,而牛顿和莱布尼茨通过无穷级数的应用认识到这一方法的普遍性.

如果费马当时认识到这一点,那么微积分的发明权就属于费马了.在数学史上,拉格朗日、拉普拉斯和傅立叶都曾称"费马是微积分的真正发明者".但泊松正确地指出,费马不应当享有这一荣誉.

但是,肯定地,除了巴罗以外,没有任何数学家像费马这样接近于微积分的发明了.

12.2.5 巴罗的贡献

另一个对微积分作出预言的是巴罗(I. Barrow).他于1630年生于伦敦,毕业于剑桥大学.他在物理、数学、天文和神学方面都有造诣.他也是当时研究古希腊数学的著名学者,他翻译了欧几里得的《几何原本》.他是第一个担任剑桥大学卢卡斯讲座教授的人.牛顿是他的学生.1669年,他辞去了他的教授席位,并赞助牛顿取得此席位.1673年被任命为剑桥三一学院院长.1677年逝世于剑桥.

巴罗最重要的著作是他的《光学和几何学讲义》.在这本书中我们能够找到非常接近近代微分过程的步骤.在本质上他已经用了今天教科书中所用的微分三角形的概念.下面我们来介绍巴罗的方法.

现在要求曲线 $f(x,y)=0$ 在点 $P(x,y)$ 处的切线.从图12-11可看出,只需定出 T 的位置.如何定 T 的位置呢?办法是作无限小的曲线三角形 PQR(暂时仍记为 ΔPQR).这时 ΔPTM 与 ΔPQR 是很接近于相似的.巴罗认为,当小三角形越来越小时,它们就是相似的.令 $QR=e$,$RP=a$.因而有

$$\frac{TM}{PM}=\frac{e}{a}\Rightarrow TM=\frac{e}{a}PM=\frac{e}{a}y.$$

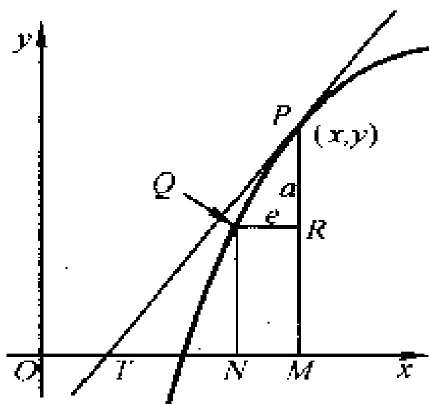


图12-11

这样一来,问题化为求 $\frac{e}{a}$.如何求 $\frac{e}{a}$ 呢?

设 P 的坐标为 (x,y),则 Q 的坐标就是 $(x-e,y-a)$.把这些值代入曲线的方程,并让 e 和 a 的二次幂和高次幂都等于0,我们就能得到比 $\frac{e}{a}$.由此,我们有

$$OT = OM - TM = OM - PM\frac{QR}{RP} = x - \frac{e}{a}y.$$

知道了 OT,就可定出点 T 的位置,连接 P 与 T 就可以确定出切线.

用现代的语言说,其核心思想是用微分三角形定 $\frac{e}{a}$,就是定切线的斜率.

例　求曲线 $x^3 + y^3 = r^3$ 在点 (x,y) 处的切线.

解　我们用巴罗的方法来求解.

把 $x - e, y - a$ 代入方程,得到

$$(x - e)^3 + (y - a)^3 = r^3$$

或

$$x^3 - 3x^2e + 3xe^2 - e^3 + y^3 - 3y^2a + 3ya^2 - a^3 = r^3.$$

让 e 和 a 的二次幂和高次幂都等于0,并利用 $x^3 + y^3 = r^3$,上式就化为

$$3x^2e + 3y^2a = 0.$$

由此我们得到

$$\frac{a}{e} = -\frac{x^2}{y^2}.$$

下面就不难求出切线了:

$$OT = x - \frac{e}{a}y = x + \frac{y^3}{x^2}.$$

从上面的求法我们看出,巴罗求切线的方法非常接近于微分学中所采用的方法.字母 a 和 e 相当于我们的符号 Δy 和 Δx.这是费马方法的进一步发展,在费马那里只用了一个无穷小量 e.并且,巴罗方法可以更方便地应用于隐函数.

尽管巴罗的方法比费马的方法更接近于求导运算,但是还不能认为它已含有我们的符号 Δy 和 Δx.由于他们没有明确的极限概念,所以他们的论证在逻辑上缺乏严格性.

除了上面的例子外,巴罗还作了以下曲线的切线:

1) $x^2(x^2+y^2)=r^2y^2$(卡铂曲线);

2) $x^3+y^3=rxy$(笛卡儿叶形线);

3) $y=(r-x)\tan(\pi x/2r)$(割圆曲线);

4) $y=r\tan(\pi x/2r)$(正切曲线).

特别有趣、特别重要的是,巴罗在《光学和几何学讲义》的第十讲和第十一讲把作曲线的切线与曲线的求积联系了起来.这就是说,他把微分学和积分学的两个基本问题以几何对比形式联系了起来.把这两个定理翻译成现代语言,并使用现代符号,则其内容可陈述如下:

a. 如果 $y=\int_0^x z\mathrm{d}x$ 则$\frac{\mathrm{d}y}{\mathrm{d}x}=z$,

b. 如果 $z=\frac{\mathrm{d}y}{\mathrm{d}x}$,则$\int_0^x z\mathrm{d}x=y$.

巴罗的确已经走到了微积分基本定理的大门口.但在巴罗的书中,这两个定理相隔二十余个别的定理,并且没有把它们对照起来,也几乎没有使用过它们.这说明巴罗并没有从一般概念意义下理解它们.但是我们知道,只有一般概念才能阐明问题的本质,才能开拓广阔的应用道路.

12.2.6 前期史小结

我们来总结一下17世纪"微积分"方面所取得的成就,总结到牛顿和莱布尼茨出现于数学界为止.

有关于现在的积分学这个范围内的成就越来越多.这里面不仅得到了大量的关于求面积、体积、弧长、曲面面积及质心定位的结果,也认识到所有在传统上归结为求面积的这类问题之间的联系.在卡瓦列里、帕斯卡等人的著作中开始结晶出定积分概念本身.实际上那时已经算出了一系列最简单的积分,常常是几何的形式,但有时也有算术的形式,找到了把一些积分化为别的积分的种种关系式.

在微分学这个领域内,费马给出了一个统一的无穷小方法,用以解决求最大、最小值问题和作曲线的切线问题.他的研究为许多其他

数学家所继续.

最后,巴罗在这两类问题中间搭成了一座桥梁.

这样一来,这门新学科的基础已经具备,但是像现在这样的微积分还没有.正如后来莱布尼茨确切表达的:"在这样的科学成就之后,所缺少的只是引出问题的迷宫的一条线,即依照代数样式的解析计算法".

在创建微积分的过程中究竟还有多少事情要做呢?

1) 需要以一般形式建立新计算法的基本概念及其相互联系,创立一套一般的符号体系,建立计算的正规程序或算法.

2) 为这门学科重建逻辑上一致的、严格的基础.

这第一个任务就是牛顿和莱布尼茨各自独立创造的微积分学所完成的工作.至于要在一个比较严格的基础上重建这个学科的基本概念,则要等到这个学科取得了广泛应用和蓬勃发展之后才可能.这是法国伟大的分析学家 A.L.柯西(Cauchy 1789—1857)及其他 19 世纪数学家的工作.

§12.3 牛顿和莱布尼茨

17 世纪后期出现了一个崭新的数学分支——数学分析,或者微积分.它在数学领域中占据着主导地位.这种新数学的特点是,非常成功地运用了无限过程的运算,即极限运算.而其中的微分和积分这两个过程则构成了微分学和积分学的核心,并奠定了全部分析学的基础.微积分的系统发展通常归功于两位伟大的科学先驱—牛顿和莱布尼茨.这一系统发展的关键在于认识到,过去一直是分别研究的微分和积分这两个过程是彼此互逆的两个过程,由牛顿—莱布尼茨公式联系着.公正的历史评价,是不能把发明微积分这一成就归功于一两个人的偶然的和不可思议的灵感.前面已经指出,许多人,例如费马、伽利略、刻卜勒,都曾为科学中的这些具有革命性的新思想所

鼓舞,对微积分的诞生作出过贡献.牛顿的老师巴罗就曾几乎充分地认识到微分和积分间的互逆关系.现在我们来扼要地谈谈牛顿和莱布尼茨这两位微积分奠基者的主要工作.

伊萨克·牛顿(Isac Newton 1643—1727)1643年1月4日生于英格兰的乌兰索普镇.在牛顿诞生前不久他的父亲就离开了人间.1660年,17岁的牛顿进入剑桥大学.1665年毕业于该校并获得学士学位,1668年获得硕士学位.1669年他继承了巴罗的职位,同时计划出一本关于导数和级数的论著,其中包括微积分基本定理.但是这份手稿一直没有发表,到他去世之后才付印,后来以流数理论著称.牛顿把导数考虑为一种速度,称之为流数.1687年,他的主要著作《自然哲学的数学原理》(以下简称《原理》)出版.在这本名著中,牛顿证明了,天体的运动可以由运动定律(力等于动量对时间的导数)和引力定律推导出来.《原理》在物理和数学的结合方面取得了首次巨大的成就,其后被许多人继承,几乎有三百年之久.牛顿在1665—1666年写出了《曲线求积论》,而在1670年写出了题为《流数术和无穷级数方法及其对几何曲线的应用》的论文.这两部著作相当迟才出版,前者在1704年出版,后者在牛顿去世后,于1736年出版.牛顿在这两部著作中叙述了数学分析的方法.在这两部著作中和在牛顿同时代人莱布尼茨的著作中,建立和完成了无穷小量的经典分析,也就是建立和完成了微积分学.牛顿的数学分析的基本概念是力学概念的反映.连最简单的几何图形——线、角、体——都被牛顿看作是力学位移的结果.线是点运动的结果,角是它的边旋转的结果,体是表面运动的结果.牛顿认为变量是运动着的点.牛顿把任何变量都叫做流动量.至今我们还用术语"流动点"表示坐标连续变化的点,即运动着的点.因为任何运动都离不开时间,所以牛顿总是把时间作为自变量.运动的速度,对我们说来是导数,牛顿把它叫做流数,并用一个点表示,如果流动量为 x,那么流数是流动量对时间的导数,现在我们用 dx/dt 来表示.

牛顿考虑了两种类型的问题.在第一种类型的问题中,给出联系一些流动量的关系式,要求找出联系这些流动量和它们的流数的关系式.这自然等价于微分.在第二种类型的问题中,给出联系一些流动量和它们的流数的关系式,要求找出仅仅联系这些流动量的关系式.这是逆问题,等价于解微分方程.牛顿完全明白这两种运算的互逆性,解决了正问题和逆问题,并把它们的解用到大量的几何问题与力学问题上去.

牛顿还研究了函数的极大值和极小值,曲线的切线,曲线的曲率,拐点,曲线的凹凸性等问题.他还给出了对代数方程和超越方程都适用的方程实根的近似值求法,这方法现在称为牛顿法(见第十章代数方程式).

牛顿最伟大的著作是他的《原理》.在《原理》中第一次有了地球和天体主要运动现象的完整的动力学体系和完整的数学公式.这是科学史上最有影响、享誉最高的著作.在爱因斯坦的相对论出现之前,这部著作是整个物理学和天文学的基础.

牛顿是人类历史上最伟大的数学家之一.像莱布尼茨这样作出了杰出贡献的人也评价道:"在从世界开始到牛顿生活的年代的全部数学中,牛顿的工作超过一半".拉格朗日称他是历史上最有才能的人,也是最幸运的人,因为宇宙体系只能被发现一次.英国著名诗人波普(Pope)是这样来描述这位伟大科学家的:

自然和自然的规律
沉浸在一片混沌之中,
上帝说,生出牛顿,
一切都变得明朗.

但是,牛顿本人却很谦虚.他说:"我不知道世间把我看成什么人;但是对我自己来说,就像一个海边玩耍的小孩,有时找到一块比较平滑的卵石或格外漂亮的贝壳,感到高兴,而在我前面是未被发现的真理的大海".

戈特弗里德·威廉·莱布尼茨(Gottfried Wilhelm Leibniz 1646—1716)1646年6月21日出生在德国莱比锡，他的父亲是莱比锡大学的道德哲学教授.莱布尼茨15岁进入莱比锡大学，学习法律.在答辩了关于逻辑的论文之后，得到哲学学士学位.1666年他写了论文《论组合的艺术》，这就完成了他在阿尔特道夫大学的博士论文，并使他获得教授席位.1670年和1671年他写了第一篇力学论文.1672年他出差到巴黎，使他接触到数学家和自然科学家，激起了他对数学的兴趣.他自己说过，直到1672年他还基本上不懂数学.

莱布尼茨研究了巴罗的著作之后，意识到微分和积分的互逆关系.他认识到，求曲线的切线依赖于纵坐标的差值与横坐标的差值之比(当这些差值变成无限小时)，而求面积则依赖于横坐标的无限小区间上的无限窄矩形面积之和.并且这种求差与求和的运算是互逆的.莱布尼茨的微分学是把微分看作变量相邻二值的无限小的差，而积分概念则是以变量分成无穷多个微分之和的形式出现.

莱布尼茨从1684年起发表微积分论文.在1684年的《博学学报》上他发表了一篇题为《一种求极大值与极小值和切线的新方法，它也适用于分式和无理量，以及这种新方法的奇妙类型的计算》.这是历史上最早公开发表的关于微分学的文献.在这篇论文中，他简明地解释了他的微分学.文中给出了微分的定义，函数的加、减、乘、除以及乘幂的微分法则，关于二阶微分的概念，以及关于微分学对于研究极值、作切线、求曲率及拐点的应用.他所给出的微分学符号和计算导数的许多一般法则一直沿用到今天.它使得微分运算几乎是机械的，而在这以前人们还不得不对每一个个别情况采用取极限的步骤。值得庆幸的另一点是，莱布尼茨引入了一套设计得很好的，令人满意的符号.莱布尼茨的符号具有独到之处.他不但为我们提供了今天正在使用的一套非常灵巧的微分学符号，而且还在1675年引入了现代的积分符号，用拉丁字Summa(求和)的第一个字母S拉长了表示积分.但是“积分”的名称出现得比较迟，它是由J.伯努利提出的.

莱布尼茨关于积分学的第一篇论文发表于1686年.他得到的积分法有:变量替换法、分部积分法、利用部分分式求有理式的积分法等.

莱布尼茨是数学史上最伟大的符号学者.他在创造微积分的过程中,花了很多时间去选择精巧的符号.他认识到,好的符号可以精确、深刻地表达概念、方法和逻辑关系.他曾说:"要发明就得挑选恰当的符号.要做到这一点,就要用含义简明的少量符号来表达或比较忠实地描绘事物的内在的本质,从而最大限度地减少人的思维劳动."现在微积分学的基本符号基本上都是他创造的.这些优越的符号为以后分析学的发展带来了极大的方便.

顺便提到的是,莱布尼茨对中国的科学文化和哲学思想十分重视.1696年他编辑出版了《中国新事萃编》一书.在该书的序言中,他说:"中国和欧州各居世界大陆的东西两端,是人类伟大的教化和灿烂文明的集中点."他主张东西方应在文化,科学方面互相学习,平等交流.他曾写了一封长达4万字的信,专门讨论中国的哲学.信的最后谈到伏羲的符号,《易经》中的64个图形与他的二进制,他说中国许多伟大的哲学家"都曾在这64个图形中寻找过哲学的秘密……,这恰恰是二进制算术.这种算术是这位伟大的创造者(伏羲)所掌握而几千年之后由我发现的."他还送过一台他制作的计算机的复制品给康熙皇帝.

综上所述,牛顿和莱布尼茨研究微积分学的基础都达到了同一目的,但各自的方法不同.牛顿主要是从力学的概念出发,而莱布尼茨作为哲学家和几何学家对方法本身感兴趣.牛顿接近最后的结论比莱布尼茨早一些,而莱布尼茨发表自己的结论早于牛顿.

§12.4　光辉的诞生

微积分的诞生具有划时代的意义,是数学史上的分水岭和转折点.这个伟大发明产生的新数学明显地不同于从古希腊继承下来的

旧数学.旧数学是关于常量的数学,而新数学是关于变量的数学.旧数学是静态的,而新数学是动态的.旧数学与新数学的关系就像解剖学与生理学,前者研究死的躯体,而后者研究活的身体.旧数学只涉及固定的和有限的,而新数学却包含了运动,变化和无限.

关于微积分的地位,恩格斯是这样评价的:"在一切理论成就中,未必再有什么像 17 世纪下半叶微积分的发现那样被看作人类精神的最高胜利了.如果在某个地方我们看到人类精神的纯粹的和唯一的功绩,那正是在这里."

我们看到了微积分的发明远非一、二人的工作.它经历了一个漫长而曲折的思想潮流,从古代的哲学思辨和数学证明引导到 17 世纪的极其成功的富于启发性的方法.17 世纪最伟大的数学家们都参与了这项伟大的工程.他们当中有刻卜勒、笛卡儿、卡瓦列里、费马、帕斯卡、罗贝瓦尔、巴罗等.最终在牛顿和莱布尼茨手中集其大成,迸发出新方法和新观点的发明,使数学达到一个更高的水平.英国的伟大诗人雪莱曾经把人类思想史上伟大进步的形成比作雪崩的形成:

一片一片的雪花,
经过风暴的再三筛选,
积成巨大的雪团.
在阳光的激发下
形成雪崩!
思想也是这样,
一点一滴地积累在
不怕上帝的人心中,
终于迸发出伟大的真理,
在万国引起响应!

第十三章　实数理论

博学之、审问之、慎思之、明辨之、笃行之.

《中庸》

人一能之,己百之;人十能之,己千之.果能此道矣,虽愚必明,虽柔必强.

《中庸》

§13.1　第二次数学危机

13.1.1　英雄世纪

微积分诞生之后,数学迎来一次空前的繁荣时期.18世纪被称为数学史上的英雄世纪.这个时期的数学家们在几乎没有逻辑支持的前提下,勇于开拓并征服了众多的科学领域.他们把微积分应用于天文学、力学、光学、热学等各个领域,并获得了丰硕的成果.在数学本身他们又发展了微分方程的理论,无穷级数的理论,大大地扩展了数学研究的范围.

18世纪的数学家们知道他们的微积分概念是不清楚的,证明也不充分,但他们却自信他们的结果是正确的.为什么会是这样呢?部分答案是,有许多结果为经验和观测所证实,其中最突出的是天文学的预言,如哈雷彗星的再度出现.另一个原因是,那时的数学家确信,上帝数学化地设计了世界,而他们正在发现和揭示这种设计.可以说,这种信仰支撑着他们的精神和勇气,而丰硕的科学成果则养育着他们的心智,成为他们追求的精神食粮.

科学上的巨大需要战胜了逻辑上的顾忌.他们需要做的事情太

多了，他们急于去攫取新的成果．基础问题只好先放一放．正如达朗贝尔所说的："向前进，你就会产生信心！"数学史的发展也一再证明自由创造总是领先于形式化和逻辑基础．

13.1.2 第二次数学危机

大家知道，在公元前 5 世纪出现了数学基础的第一次灾难性危机，这就是无理数的诞生．这次危机的产生和解决大大地推动了数学的发展．

在微积分的发展过程中，一方面是成果丰硕，另一方面是基础的不稳固，出现了越来越多的谬论和悖论．数学的发展又遇到了深刻的令人不安的危机．由微积分的基础所引发的危机在数学史上称为第二次数学危机．

虽然在牛顿和莱布尼茨创立微积分之后的大约一百年中，很少注意到从逻辑上加强这门学科的基础，但绝不是对薄弱的基础没有人批评．一些数学家进行过长期的争论，并且，两位创立者本人对此学科的基本概念也不满意．对有缺陷的基础最强有力的批评来自一位非数学家，这就是著名的唯心主义哲学家贝克莱主教(Bishop George Berkeley, 1685 − 1753)．他坚持：微积分的发展包含了偷换假设的逻辑错误．我们以考察牛顿对现在称作微分所采用的方法，来弄明白这个特殊的批判．

牛顿在 1704 年发表了"曲线的求积"，其中他确定了 x^3 的导数(他当时称为流数)．我们把牛顿的方法意译如下：

当 x 增长为 $x + 0$ 时，幂 x^3 成为 $(x + 0)^3$，或 $x^3 + 3x^2 0 + 3x0^2 + 0^3$．

它们的增量分别为 0 和 $3x^2 0 + 3x0^2 + 0^2$．

这两个增量与 x 的增量 0 的比分别为 1 与 $3x^3 + 3x0 + 0^2$．

然后让增量消失，则它们的最后比将为 1 比 $3x^2$．从而 x^3 对 x 的变化率为 $3x^2$．

从引文中可看出，偷换假设的错误是明显的．在论证的前一部

分,假定 0 是非零的,而在论证的后一部分,它又被取为零.贝克莱说:“在我们假定增量消失时,理所当然,也得假设它的大小、表达式以及其它,由于它的存在而随之而来的一切也随之消失.”他还说:“总之,不论怎样看,牛顿的流数算法是不合逻辑的”.这就是历史上著名的《贝克莱悖论》.

这里指出,只是在对极限论作了严格的逻辑处理之后,这一困难和缺陷才得以克服.

为了使读者对当时分析中出现的谬误有所了解,我们再看看大数学家欧拉在他使用分析推理时出现的一些悖论.

把二项式定理形式地应用于$(1-x)^{-1}$,我们得到

$$\frac{1}{1-x}=(1-x)^{-1}=1+x+x^2+x^3+\cdots,$$

然后令 $x=2$,我们有

$$-1=1+2+4+8+16+\cdots.$$

这就是欧拉得到的一个不得不接受的荒谬结论.还有,把前式两边乘以 x,得

$$x+x^2+\cdots=\frac{x}{1-x}.$$

另一方面,

$$\frac{x}{x-1}=\frac{1}{1-\frac{1}{x}}=1+\frac{1}{x}+\frac{1}{x^2}+\cdots.$$

两式相加后,欧拉得到

$$\cdots+\frac{1}{x^2}+\frac{1}{x}+1+x+x^2+\cdots=0.$$

这也是一个十分荒谬的结果.

17 世纪和 18 世纪的数学家们对无穷级数不大理解,以致在分析这个领域中出现了许多悖论.再如,考虑级数

$$S=1-1+1-1+1-1+\cdots,$$

如果把级数以一种方法分组,我们有

$$S = (1-1)+(1-1)+(1-1)+\cdots = 0,$$

如果按另一种方法分组,我们有

$$S = 1-(1-1+1-1+1-1+\cdots) = 1-0 = 1.$$

L.G.格兰迪(Grandi, 1671 - 1742)说,因为0和1是等可能的,所以级数的和应为平均数1/2.这个值也能用纯形式的方法得到.事实上,

$$S = 1-(1-1+1-1+1-1+\cdots) = 1-S.$$

由此得 $2S = 1$,或 $S = 1/2$.

这样的悖论日益增多.数学家们在研究无穷级数的时候,做出许多错误的证明,并由此得到许多错误的结论.他们在有限与无限之间任意通行,他们的工作可以用伏尔泰的一句话来概括:微积分是“计算和度量一个其存在性是不可思议的事物的艺术”.

因此在18世纪结束之际,微积分和建立在微积分基础上的分析的其它分支的逻辑处于一种完全混乱的状态之中.事实上,可以说微积分在基础方面的状况比17世纪更差.数学巨匠,尤其是欧拉和拉格朗日给出了不正确的逻辑基础.因为他们是权威,所以他们的错误就被其他数学家不加批判地接受了,甚至作了进一步的发展.

进入19世纪,数学陷入更加矛盾的境地.虽然它在描述和预测物理现象方面所取得的成就远远超出人们的预料,但是大量的数学结构没有逻辑基础,因此不能保证数学是正确无误的.

历史要求给微积分以严格的基础.

13.1.3　柯西的功绩

第一个为补救第二次数学危机提出真正有见地的意见的是达朗贝尔.他在1754年指出,必须用可靠的理论去代替当时使用的粗糙的极限理论.但是他本人未能提供这样的理论.最早使微积分严谨化的是拉格朗日.为了避免使用无穷小推理和当时还不明确的极限概念,拉格朗日曾试图把整个微积分建立在泰勒展式的基础上.但是,

这样一来，考虑的函数范围太窄了，而且不用极限概念也无法讨论无穷级数的收敛问题．所以，拉格朗日的以幂级数为工具的代数方法也未能解决微积分的奠基问题．

到了19世纪，出现了一批杰出的数学家，他们积极为微积分学的奠基工作而努力．首先要提到的是捷克的哲学家和数学家波尔查诺(B. Bolzano，1781—1848)．他开始将严格的论证引入到数学分析中．1816年，他在二项展开公式的证明中，明确地提出了级数收敛的概念，同时对极限、连续和变量有了较深入的理解．特别是，他曾写出《无穷的悖论》一书，书中包含许多真知灼见．可惜，在他去世两年后该书(1850年)才得以出版．

分析学的奠基人，公认是法国多产的数学家柯西(A. L. Cauchy，1789—1857)．柯西在数学分析和置换群理论方面做了开拓性的工作，是最伟大的近代数学家之一．他在1821—1823年间出版的《分析教程》和《无穷小计算讲义》是数学史上划时代的著作．在那里他给出了数学分析一系列基本概念的精确定义．例如，他给出了精确的极限定义，然后用极限定义连续性、导数、微分、定积分和无穷级数的收敛性．这些定义基本上就是今天我们微积分课本中使用的定义，不过现在写得更加严格一点．

13.1.4 魏尔斯特拉斯的规划

对分析基础作更深一步的理解的要求发生的1874年．那时德国数学家魏尔斯特拉斯(K. T. W. Weirstrass，1815—1897)构造了一个没有导数的函数，即构造了一条处处没有切线的连续曲线．这与直观信念是有矛盾的．这对在分析学中运用几何直观是一场大风暴．

连续性和可微性是分析学的基本概念，从微积分诞生起一直是分析研究的主要对象．但是数学家们对它们的认识一直是模糊不清的．甚至有的数学家还证明过，任何函数在所有的连续点上都有导数．魏尔斯特拉斯的函数几乎使所有的数学家都感到震惊．埃尔米特在1893年5月20日给斯蒂杰斯的信中写道："我简直惊恐万状，不愿

面对这一不幸的现实，没有导数的连续函数!”

极限概念、连续性、可微性和收敛性对实数系的依赖比当时人们想象的要深奥得多.黎曼发现，柯西没有必要把他的定积分限制于连续函数.黎曼证明了，被积函数不连续，其定积分也可能存在.黎曼还造出一个函数，当变量取无理值时它是连续的；当变量取有理值时它是不连续的.这些例子使我们越来越明白，在为分析建立一个完善的基础方面，还需要再深挖一步：我们需要理解实数系的更深刻的性质.

这个任务落在了魏尔斯特拉斯身上.魏尔斯特拉斯提出一个规划:1）逻辑地构造实数系;2）从实数系出发去定义极限概念、连续性、可微性、收敛和发散.

这个规划称为分析的算术化.任务是繁重而困难的，但在接近19世纪末的时候，这个规划终于完成了.

魏尔斯特拉斯的努力终于使分析从完全依靠运动学、直觉理解和几何概念中解放了出来.魏尔斯特拉斯规划的成功产生了深远的影响.主要表现在以下几点：

1）既然分析能从实数系导出，所以，如果实数系是相容的，那么全部分析是相容的.

2）欧氏几何通过笛卡儿坐标系也能奠基于实数系上.所以，如果实数系是相容的，那么欧氏几何是相容的，几何学的其它分支也是相容的.

3）实数系可用来解释代数的许多分支，所以许多代数的相容性也依赖于实数系的相容性.

由此得到，如果实数系是相容的，那么大部分数学就是相容的.

魏尔斯特拉斯规划的第二部分是由引进精确的“$\varepsilon-\delta$”语言而完成的.这一语言给出极限的准确描述，消除了历史上各种模糊的用语，诸如“最终比”、“无限地趋近于”等等.这样一来，分析中的所有基本概念都可以通过实数和它们的基本运算和关系精确地表述出

来.

魏尔斯特拉斯规划的第一部分就是实数系的建立,下面我们将详细讨论.

总之,第二次数学危机的核心是微积分的基础不稳固.柯西的贡献在于,将微积分建立在极限论的基础上.遗留的问题是,任何实数列的极限存在吗?魏尔斯特拉斯的贡献在于,先逻辑地构造实数系.因而,建立分析基础的逻辑顺序是实数系 — 极限论 — 微积分.

§13.2　实数集合的基本性质

13.2.1　从有理数谈起

我们用 Q 表示全体有理数的集合.Q 有一个重要性质,就是"稠密性",我们在第一讲就提到过.对于任意两个有理数 r_1, r_2,不管它们相距多近,即不管 $|r_1 - r_2|$ 多么小,它们之间总有另一个有理数 r_3,例如可取 $r_3 = \frac{1}{2}(r_1 + r_2)$.在 r_1 和 r_3 之间,r_2 和 r_3 之间,还有另外的有理数.如此类推,可知,在 r_1 和 r_2 之间存在无穷多个有理数.

在数轴上,以有理数为坐标的点叫作有理点.数集 Q 与数轴上的有理点是一一对应的,所以我们也可以用 Q 表示有理点集.分析学上习惯地将"点"和"数"混用不分.

Q 的稠密性的另一种表示法是,给定一个有理数 r,总存在有理点列 $\{r_n\}$ 以 r 为极限,即 $\lim\limits_{n\to\infty} r_n = r$.这表明有理点列间的距离可以缩为 0.由此可知,有理点之间不可能存在大于 0 的最小间距.也就是,任何有理数 r 附近不存在与 r 最近而又异于 r 的有理点.

有理数是我们日常生活和科学技术中用于测量的数.它的稠密性保证了我们的测量可以达到任意高的精确度.就实用而言,有理数是完全够用的.此外,有理数系对四则运算是封闭的,它是我们遇到

的第一个比较完美的数系.

但是,有理数系仍然存在严重的缺陷.

首先,有理点并没有填满整个数轴.我们早就证明了$\sqrt{2}$就不是有理点,而且知道,有理点间的空隙非常之多.从几何上看,这是不完美的.如果不把它们填补起来,就连平面几何中"截取交点"这件事都会遇到麻烦.从代数上看,有理数系对开方运算不封闭,有理数的开方可能不再是有理数.这个问题不解决,进一步的代数运算将遇到麻烦.

其次,当我们从变量的角度考察问题时,就会发现,有理数系在极限运算下不封闭.即,由有理数组成的序列,其极限可能不再是有理数.有理数的这种不完备性,是一个本质上的缺陷.它使得有理数系不能成为微积分学的立论的基础.

例 考察有理数序列$\{S_n\}$,其中

$$S_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} \quad (n = 1,2\cdots),$$

这个序列是读者熟悉的,它的极限是数 e(它的存在性在后面给出).数 e 是数学分析中使用最广泛的常数之一.我们曾指出,这个数是超越数,但是没有证明.现在我们来证明它不是有理数.对于任何$n > m$,我们有

$$\begin{aligned} S_n &= S_m + \frac{1}{(m+1)!} + \frac{1}{(m+2)!} + \cdots + \frac{1}{n!} \\ &\leqslant S_m + \frac{1}{(m+1)!}\left[1 + \frac{1}{m+2} + \frac{1}{(m+2)(m+3)} + \cdots\right] \\ &\leqslant S_m + \frac{1}{(m+1)!}\left[1 + \frac{1}{m+1} + \frac{1}{(m+1)^2} + \cdots\right] \\ &= S_m + \frac{1}{(m+1)!} \cdot \frac{1}{1 - \dfrac{1}{m+1}} = S_m + \frac{1}{m} \cdot \frac{1}{m!}, \end{aligned}$$

因此,当$n > m$时,

$$S_m < S_n \leqslant S_m + \frac{1}{m} \cdot \frac{1}{m!}.$$

令 n 无限增大,而 m 保持不变,我们得到,

$$S_m < \mathrm{e} \leqslant S_m + \frac{1}{m} \cdot \frac{1}{m!}.$$

因此,e同 S_m 的差最大是 $\frac{1}{m} \cdot \frac{1}{m!}$. 注意到,$m!$随 m 增大而极其迅速地增大,所以对于适当的 m,数 S_m 已经是e的很好的近似值. 例如,S_{10} 同e的差小于 10^{-7}. 用这种方法我们可以求出 $\mathrm{e} = 2.718\,281\cdots$.

现在我们来证明e是无理数. 假定e是有理数,并将它写为 $\mathrm{e} = \frac{p}{m}$,$m > 1$. 因为e位于2与3之间,所以它不是整数. 根据前面的论证,我们有,

$$S_m < \frac{p}{m} \leqslant S_m + \frac{1}{m} \cdot \frac{1}{m!}.$$

上式两端乘以 $m!$,得到

$$m!S_m < p(m-1)! \leqslant m!S_m + \frac{1}{m} < m!S_m + 1,$$

但是,

$$m!S_m = m! + m! + \frac{m!}{2!} + \frac{m!}{3!} + \cdots + \frac{m!}{m!}$$

是一个整数,因为右端和式中每一项都是整数. 这样一来,如果e是有理数,则整数 $p(m-1)!$将位于两个相继的整数之间,而这是不可能的. 这样我们就证明了e是无理数.

这个例子说明,有理数系在极限运算下不是封闭的.

总之,如果不把有理数域进行扩充,那么从几何上看,数轴是不完备的,存在许多空隙;从代数上看,开方运算不封闭;从分析上看,极限运算不封闭.

正是有理数的这些缺陷,引出了康托尔、戴德金、魏尔斯特拉斯等人对无理数本质的深刻研究,并奠定了实数的构造理论.

关于实数的构造,已有三种不同的方法,也就是有三派理论,即戴德金的"分划",康托尔－海涅的"基本序列"和魏尔斯特拉斯的"有界单调序列".这三种构造方法有一个共同点:都是利用有理数的某些集合来定义无理数.并且,这三种定义在逻辑上是等价的.一旦证明了它们的等价性,我们就可以从任一定义出发去定义实数.我们采用戴德金的"分划"来定义实数.因为这种构造法的直观性强,是多数教科书采用的方法.但是必需指出,无论哪种构造都采用了公理化的方法,都具有一定的抽象性,读起来要花一些心思.

13.2.2 戴德金分划

戴德金引出分划概念的想法,是从如何定义$\sqrt{2}$得来的.我们知道,$\sqrt{2} \notin \mathbf{Q}$.根据有理数的稠密性,不难找到有理点列 r_n, r_n',满足条件

$$\lim_{n\to\infty}(r_n' - r_n) = 0, \quad r_n < \sqrt{2} < r_n'.$$

从直观上看,$\sqrt{2}$恰好是两个有理点列$\{r_n\}$和$\{r_n'\}$的分划点.但是,我们还不知道$\sqrt{2}$为何物,又如何选取有理点列 r_n 使它与$\sqrt{2}$作比较呢?这个问题倒不难回答,因为我们总可以选取 $r_n \in \mathbf{Q}$,使得 $r_n^2 < 2$.同理,可以选取 $r_n' \in \mathbf{Q}$, $r_n'^2 > 2$.

满足上面条件的点列显然是非常多的.为了避免特殊性和不确定性,我们把点集 $\mathbf{Q}$ 分为两类 A 和 A'.其中

A 类:一切使 $r^2 < 2$ 的正有理数 r,零,及一切负有理数.

A' 类:一切使 $r^2 > 2$ 的正有理数 r.

这个分划具有这样三条性质:1) 集 A 和集 A' 都是非空的;2) $A \cup A' = \mathbf{Q}$,即没有漏掉一个有理数;3) 集 A 中的每一个数都小于集 A' 中的每一个数.用一句话来描述分划的这三条性质就是:不空,不漏,不乱.

下面我们引入戴德金关于分划的定义.

定义 把全体有理数的集合分成两个集合 A 和 A',满足下面

三个条件：

1）集合 A 和 A' 都是非空的（不空）；

2）每一个有理数在而且只在 A 与 A' 两个集合的一个之中（不漏）；

3）集合 A 中的每一个数 a 都小于集合 A' 中的每一个数 a'（不乱）.

集合 A 叫作分划的下类，集合 A' 叫作分划的上类. 用 $A|A'$ 表示这一分划.

由分划的定义可知，凡小于下类中的数 a 的有理数也属于下类. 同样，凡大于上类中数 a' 的有理数也属于上类.

例1　把 A 定义为一切满足不等式 $a<1$ 的有理数 a 的集合，而把 $a'\geqslant 1$ 的一切有理数 a' 归入集合 A'.

不难验证，我们确实得到了一个满足定义的分划，即满足分划的三个条件. 特别地，数 1 属于集合 A'，并且它是 A' 中的最小的数. 另一方面，A 类中没有最大的数. 因为，不论我们在 A 中取怎样的数 a，我们总可以在 a 和 1 之间找出一个有理数 a_1，$a_1>a$，且属于 A 类.

例2　把 A 定义为一切满足不等式 $a\leqslant 1$ 的有理数 a 的集合，而把 $a'>1$ 的一切有理数 a' 归入集合 A'.

这也是一个分划，并且在上类中没有最小数，而在下类中有最大数.

例3　由 $\sqrt{2}$ 所产生的分划.

定义如前，这也是一个分划，并且在上类中没有最小数，在下类中没有最大数.

命题　不存在 $\mathbf{Q}$ 的这样的分划 $A|A'$，使 A 中有最大数，A' 中有最小数（自证）.

可见，分划只能有三种类型：

1）在上类中没有最小数，而在下类中有最大数 r；

2）在上类中有最小数 r，而在下类中没有最大数；

3）在上类中没有最小数，在下类中也没有最大数.

在前两种情形下，我们说，这个分划由有理数产生，或者说，这个分划定义了有理数 r. 在例 1 和例 2 中这样的数 r 是 1，它是两类集合的界数.

在第三种情形下界数不存在，分划不能定义任何有理数. 我们现在需要引进一类新数 —— 无理数. 下面给出它的定义.

定义　任何属于类型 3）的分划定义了一个无理数 α.

这个新数 α 就代表着缺少了的界数，我们把它插在类 A 和类 A' 之间. 例 3 就是 $\sqrt{2}$ 的定义.

有理数与无理数统称为实数. 实数的概念不仅是数学分析的一个基本概念，而且也是整个数学的基本概念.

对于每一个有理数 r，存在两个定义它的分划. 在两个分划中都是数 $a < r$ 归入下类，数 $a' > r$ 属于上类，而数 r 本身可以在上类，也可以在下类. 为了确定起见，我们规定：凡说到有理数的分划时，总是把这个数归入上类. 这样一来，下类 A 中没有最大数.

13.2.3　实数的性质

有了实数的严格定义之后，我们的下一个任务就是在此定义的基础上研究实数的性质. 实数有哪些基本性质呢？我们说，实数有三种基本性质.

1）实数的有序性. 这就是说，实数集是一个有序集合，任何两个实数可以比较大小. 这样一来有理数系的有序性扩充到了实数系. 我们知道，这一性质不能扩充到复数系. 这说明，有序性是实数系所具有的一个重要性质，而不是任何数系都具有的性质.

2）实数的连续性，或实数的完备性. $\sqrt{2}$ 的例子告诉我们在有理数之间有空隙，它们没有把数轴填满. 实数之间有没有空隙呢？给定一个满足上面三条性质的实数的分划，这个分划还会有空隙吗？例如，给定一个正实数，它的平方根是实数吗？会不会出现正实数的平方根超出实数域的例外情形呢？再如，给定一个实数序列，这个序列

的极限一定是实数吗?会不会又像有理数那样,出现实数序列的极限不是实数呢?下面将证明不会出现这种情况,实数系是完备的.实数系的这个性质属于拓扑性质.

3) 实数的代数结构.我们都很熟悉,有理数系可进行四则运算,并且这些运算遵从分配律、交换律、结合律等运算规律.这些运算规律对实数仍成立,但需要给予严格证明.

下面我们分别来研究这三条基本性质.

13.2.4　实数集合的有序化

现在我们利用有理数集的有序性来建立实数集合的有序性.这需要给出两个实数相等的定义,并在实数集合中给出大小的概念.

定义　由分划 $A|A'$ 和 $B|B'$ 分别定义的两个实数 α 和 β,当且仅当这两个分划相同时才相等.

按照上面的规定,只要分划的两个下类 A 和 B 相同即可,因为这时两个上类 A' 和 B' 也必相同.这个定义对 α 和 β 是有理数自然也成立.

现在我们来建立关于实数的"大于"或"小于"的概念.对于有理数来说,这个概念从中学课本中已经知道了.对于有理数 r 与无理数 α 来说,在分划的定义中实际上已经包含了.如果 α 是由分划 $A|A'$ 定义的,我们就算作 α 大于所有 A 类中的数,而小于所有 A' 类中的数;r 一定在 A 中或在 A' 中.

现在设有两个无理数 α 和 β,α 由分划 $A|A'$ 确定,β 由分划 $B|B'$ 确定.我们把具有较大的下类的那个数算作较大的.确切地说,

定义　若 A 类完全包含 B 类,且不与 B 类相同,则称 $\alpha > \beta$,或 $\beta < \alpha$.

不难验证,这个定义在 α 和 β 两数中有一个数是有理数,或两个都是有理数时也成立.

注意,我们这里是用集合的包含关系来定义实数的大小的,所以要换一种思想方法来考虑问题.

由上面的定义立刻可得到下面的命题.

定理1 任何两个实数和之间必有下列三种关系之一:

$$\alpha = \beta, \alpha > \beta, \alpha < \beta.$$

其次,$\alpha > \beta, \beta > \gamma \Rightarrow \alpha > \gamma$.

这个定理建立了实数集合的序关系:任何两个实数都可以比较大小,并且大小关系可以传递.

为了以后证明的方便,我们建立下面的引理.

引理1 设 α, β 是两个任意的实数.若 $\alpha > \beta$,则总可以找到有理数 r,使之介于 α 和 β 之间:$\beta < r < \alpha$.

证 因为 $\alpha > \beta$,所以 α 定义的分划的下类 A 完全包含了定义 β 的下类 B,并且 A 与 B 不相同.这样一来,存在有理数 r 满足:$r \in A, r \notin B$.因而 r 属于 B'.从而 $\beta \leqslant r < \alpha$(等式只有在 β 是有理数时才成立).但因在 A 中没有最大数,所以我们一定可以取到有理数 r 使得

$$\beta < r < \alpha.$$

证毕.

引理2 设 α, β 是两个给定的实数.如果对无论怎样小的有理数 $e > 0$,总能使 α 与 β 夹在两个同样的有理数中间:

$$s \leqslant \alpha \leqslant s', s \leqslant \beta \leqslant s'$$

其中 $s' - s < e$,则数 α 与 β 一定相等.

证 反证法.假定 $\alpha \neq \beta$.不失一般性,可假定 $\alpha > \beta$.由引理1,在数 α 与 β 之间可以插入两个有理数 r 与 $r' > r$:

$$\beta < r < r' < \alpha.$$

对任意满足引理条件的 s 与 s',易见,

$$s < r < r' < s' \Rightarrow s' - s > r' - r > 0,$$

可见差 $s' - s$ 不能任意小,与引理的条件相矛盾.证毕.

13.2.5 实数集合的连续性

现在我们转过来讨论全部实数集合的一个极为重要的性质,这个性质把它和有理数集合从本质上区别开来.这个性质就是实数的

完备性，或连续性. 在考虑有理数集合的分划时，有时有这样一种分划存在，在这种分划中没有产生分划的界数. 正是在有理数的集合中留有这种空隙，使有理数集失去了完备性，而为引进新的数——无理数提供了根据. 现在我们来讨论全部实数集合中的分划.

把全体实数的集合分成两个集合 A 和 A'，满足下面三个条件：

1）集合 A 和 A' 都是非空的(不空)；

2）每一个实数在而且只在 A 与 A' 两个集合的一个之中(不漏)；

3）集合 A 中的每一个数 a 都小于集合 A' 中的每一个数 a'(不乱).

集合 A 叫作分划的下类，集合 A' 叫作分划的上类. 仍用 $A|A'$ 表示这一分划.

于是出现这样一个问题：对这种分划来说，在实数集合中总能找到产生分划的界数，还是在实数集合中仍有空隙?

下面的定理指出，空隙不复存在，实数集合是完备的.

定理2(戴德金) 对实数集合的任何分划 $\mathbf{A}|\mathbf{A}'$，都存在产生这个分划的实数 α，这个数 α 或者是下类 $\mathbf{A}$ 中的最大数，或者是上类 $\mathbf{A}'$ 中的最小数.

注 实数集合的这个性质叫作实数集合的完备性，或连续性.

证 用 A 表示所有属于 $\mathbf{A}$ 的有理数集合，用 A' 表示所有属于 $\mathbf{A}'$ 的有理数集合. 易见，A 和 A' 作成全部有理数集合的一个分划.

分划 $A|A'$ 定义了某个实数 α. 它应该落在 $\mathbf{A}$，$\mathbf{A}'$ 两类之一. 假定它落在下类 $\mathbf{A}$ 中，这时我们来证明，α 是 $\mathbf{A}$ 中的最大数. 假如不然，则可以找到另一个大于 α 的数 $\alpha_1 \in \mathbf{A}$. 根据引理1，我们可在 α 和 α_1 之间插入一个有理数 r：$\alpha < r < \alpha_1$. r 属于 $\mathbf{A}$ 类，因而也属于 A 类. 于是我们就得出了矛盾：在定义数 α 的分划的下类中会有有理数比 α 这个数更大！这样我们就证明了，α 是 $\mathbf{A}$ 中的最大数.

同理可证明，如果 α 落在上类 $\mathbf{A}'$ 中，则 α 是 $\mathbf{A}$ 中的最小数.

不会出现 A 类中有最大数，而 A' 类中有最小数的情形.

α 就是产生分划 $A|A'$ 的数. 证毕.

这个定理是实数理论的第一个重要定理，其它定理将以此定理为基础，我们把它叫作戴德金基本定理. 直观讲，在数轴上随便砍一刀，不会落在空隙中，一定会落在某一实数上. 数轴是连绵不断的. 这就是连续性与完全性的直观含义. 数轴上的点与实数集合可以建立一一对应，实现了几何与代数的完全统一.

13.2.6 确界存在定理

现在我们利用戴德金定理来建立在近代分析学上起重要作用的一些基本概念. 这些概念对讨论实数的算术运算也是必需的.

设 E 是一个实数集合. 如果存在数 M，使得对所有的 $x \in E$，都有 $|x| < M$，或 $-M < x < M$，我们就说，集合 E 是有界集.

如果 E 不满足上述条件，即对任意的正数 M，不管它多大，总有 $x_0 \in E$，使得 $|x_0| > M$，我们就称 E 为无界集.

对于集合 E 来说，如果存在数 K（或 k），使得对所有的数 $x \in E$ 都有 $x \leqslant K$（或 $x \geqslant k$），我们就称集合 E 有上界（或者有下界）. 数 K（或 k）称为集合 E 的一个上界（或下界）.

有界集指的是，同时有上界和下界. 全体实数的集合既无上界也无下界. 正实数集合有下界，但没有上界. 当 a, b 是有限数时，闭区间 $[a, b]$ 和开区间 (a, b) 都是有界集的例子.

下面我们引进一个在分析中十分重要的概念，即确界的概念.

考虑数集 A. 如果常数 M（或 m）具有下列性质：

1）对于所有的 $x \in A$，$x \leqslant M$（或者 $x \geqslant m$）；

2）不管 $\varepsilon > 0$ 是多么小的数，总可以找到 $x_0 \in A$，使得 $M - \varepsilon < x_0$（或者 $x_0 < m + \varepsilon$），那么就称数 M（或 m）是数集 A 的上确界（或者下确界）.

A 的上确界记作

$$M = \sup A = \sup_{x \in A} x.$$

而下确界记作

$$m = \inf A = \inf_{x \in A} x.$$

其中 sup 是拉丁文 supremum(最高的)的缩写;inf 是 infimum(最低的)的缩写.

上确界是上界中最小的,下确界是下界中最大的.

定理 3 如果 $H = \{x\}$ 是有上(下)界的集合,则它一定有上(下)确界.

证 我们就上确界来进行证明,下确界的证明完全一样.考虑两种情况:

1. 若在集合 H 中可以找到最大的数 $\bar{x}$,于是集合 H 中的所有的数都满足不等式 $x \leqslant \bar{x}$,即 $\bar{x}$ 是 H 的上界.另一方面,$\bar{x}$ 属于 H,因此,对于 H 的任何上界 M 都有 $\bar{x} \leqslant M$.从而 $\bar{x}$ 是上界中最小的,即它是 A 的上确界.

2. 现在设集合 H 中没有最大数.我们作下面的分划.把 H 的一切上界归入类 A',把其余的一切实数归入下类 A.在这个分法之下,集合 H 中的所有数都落在 A 类,因为由假设,H 中任何一数都不是最大的.易见,A 和 A' 两类都是不空的.这个分法实际上是一个分划,因为全部实数都分布在这两类中,并且 A' 类中每一个数都大于 A 类中任何数.换言之,分划满足不空、不乱和不漏这三个条件.根据定理 2,一定有一个作成这个分划的数 α 存在.属于 A 类的一切数 x 不能超过这个界数 α,因而 α 是一个上界,即 α 属于 A' 类而且是 A' 类的最小数.可见,α 是集合 H 的上确界.证毕.

显然,闭区间 $[a,b]$ 和开区间 (a,b) 的上确界是数 b.对于闭区间来说,上确界属于它,而对于开区间来说,上确界不属于它.

无上界集 H 记作

$$\sup H = \sup_{x \in H} x = +\infty.$$

无下界集 H 记作

$$\inf H = \inf_{x \in H} x = -\infty.$$

确界存在定理是一个强有力的定理，下面我们将不断看到它的威力：许多极限的存在性就是靠这个定理建立的．例如，数 e 的存在性．

习　题

1．求下列函数的上、下确界：

1) $f(x) = x^2$ 在 $(-2,5)$ 内；　2) $f(x) = [x]$ 在 $(0,2)$ 内．

2．设 $f(x), g(x)$ 都定义在区间 $[a,b]$ 内，证明

$$\inf(f(x) + g(x)) \geqslant \inf f(x) + \inf g(x),$$

$$\sup(f(x) + g(x)) \leqslant \sup f(x) + \sup g(x).$$

§ 13.3　实数的四则运算

有了实数定义之后，我们来讨论如何对实数进行运算，以及这些运算所遵循的运算法则．首先建立实数的四则运算，然后建立实数的乘方与开方运算．如何定义这些运算呢？必须回到实数的定义．实数是用分划定义的，实数的运算也必然用分划来定义．

在四则运算中，加法运算和乘法运算是基本的，减法和除法是加法和乘法的逆运算．我们将把重点放在加法运算和乘法运算上．其次，把加法的定义弄清楚了乘法的定义自然也就清楚了．

13.3.1　实数和的定义

设有两个实数 α 和 β．我们考虑满足不等式

$$a < \alpha < a', b < \beta < b' \tag{1}$$

的有理数 a, a' 与 b, b'．

我们定义数 α 与 β 的和是这样的实数，它介于所有形如 $a+b$ 之和与所有形如 $a'+b'$ 之和的中间：

$$a+b<\gamma<a'+b'. \tag{2}$$

这是我们利用实数定义给出的第一个运算——加法运算.我们必须对定义作出说明：

1）这个定义是合理的；对任何两个实数，和存在且唯一.

2）当 α,β 是有理数时，这样定义的和与我们过去定义的和一致.

先证明定义的合理性：γ 是存在且唯一的.

存在性.我们必须找出一个满足定义的实数.办法有二，一是作分划，用戴德金定理；一是用确界定理.我们用确界定理.

考虑一切可能的和 $a+b$ 的集合.这个集合是有上界的.易见，任何一个形如 $a'+b'$ 的和就是这个集合的一个上界.令这个集合的上确界是

$$\gamma=\sup\{a+b\}.$$

于是 $a+b\leqslant\gamma$，并且 $\gamma\leqslant a'+b'$.

因为 a,b,a',b' 是满足条件(1)的有理数，所以，总能加大 a,b 两数，减少 a',b' 两数，而使刚才得到的不等式中去掉等式，即使(2)成立.由此可见，数 γ 满足和的定义.

唯一性.我们证明，上面定义的和 γ 是唯一确定的.

设 e 是任意小的正有理数.取满足(1)的有理数 a,a',b,b'，使得

$$a'-a<e,b'-b<e.$$

由此得，

$$(a'+b')-(a+b)=(a'-a)+(b'-b)<2e.$$

这就是说，这个差数可以任意地小.由 §2 引理 2，在 $a+b$ 和 $a'+b'$ 间存在唯一的一个数.

这样一来，和 γ 是存在且唯一的.

最后，我们注意，如果 α 与 β 是两个有理数，则它们的通常的和 $\gamma = \alpha + \beta$ 显然满足不等式(2). 因此，上面给出的两个实数之和的一般定义与两个有理数之和的旧定义是一致的.

对于实数来说，加法的所有性质仍然保持：

1) $\alpha + \beta = \beta + \alpha$;

2) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;

3) $\alpha + 0 = \alpha$;

4) $\alpha > \beta \Rightarrow \alpha + \gamma > \beta + \gamma$.

根据前面所给出的和的定义以及有理数的一些熟知的性质，不难证明上面的四条性质，我们把它们的证明留给读者.

13.3.2 对称数

为了定义减法，我们需要引进对称数的概念. 我们先证明：

每一个实数 α 都存在一个满足条件 $\alpha + (-\alpha) = 0$ 的数 $(-\alpha)$，称它为 α 的对称数.

这时只需要考虑无理数的情形.

假定数 α 由分划 $A|A'$ 来确定，我们用下面的方法来定义数 $-\alpha$. 把一切有理数 $-a'(a' \in A')$ 归入数 $-\alpha$ 的下类 $\bar{A}$，而把有理数 $-a(a \in A)$ 归入这个数的上类 $\bar{A}'$. 不难看出，所作的分法是一个分划，满足不空、不漏和不乱三个条件，从而定义了一个实数. 用 $-\alpha$ 表示这个实数.

我们来验证它满足上面所说的条件. 由 $-\alpha$ 的定义，我们知道，

$$a - a' < \alpha + (-\alpha) < a' - a.$$

另一方面，

$$a - a' < 0 < a' - a,$$

而 $a' - a$ 可以任意地小，由 § 2 引理 2，满足这种条件的数是唯一的，所以

$$\alpha + (-\alpha) = 0.$$

这就是要证明的结论. 由对称数的定义，不难看出下面的性质：

$$(-\alpha) = \alpha, -(\alpha+\beta) = (-\alpha)+(-\beta).$$

13.3.3 实数减法的定义

利用对称数的概念可以把减法定义为加法的逆运算.

实数 α 与 β 的差定义为满足条件

$$\gamma+\beta=\alpha(\Leftrightarrow\beta+\gamma=\alpha) \tag{3}$$

的数 γ,记为 $\gamma=\alpha-\beta$.

和加法一样,我们需要指出,这样的数 γ 是存在且唯一的.

先证存在性.事实上,令

$$\gamma=\alpha+(-\beta),$$

则 γ 满足(3),因为

$$\begin{aligned}\gamma+\beta&=[\alpha+(-\beta)]+\beta=\alpha+[(-\beta)+\beta]\\&=\alpha+[\beta+(-\beta)]=\alpha+0=\alpha.\end{aligned}$$

唯一性也是明显的.若另有 γ' 满足(3),也必有

$$\gamma'=\alpha+(-\beta).$$

事实上,若

$$\gamma'+\beta=\alpha,$$

则 $$\gamma'+\beta+(-\beta)=\alpha+(-\beta)\Leftrightarrow\gamma'=\alpha+(-\beta),$$

这样一来,

$$\gamma=\alpha+(-\beta)=\alpha-\beta.$$

13.3.4 实数的绝对值

实数的绝对值的概念与对称数有密切联系.由对称数的构造可知,当 $\alpha>0$ 时,$-\alpha<0$,而由 $\alpha<0$ 可推出 $-\alpha>0$.换言之,只要 $\alpha\neq0$,就知道 α 与 $-\alpha$ 两数中一定有一个大于0,这就是 α 与 $-\alpha$ 的绝对值.

定义 实数 α 的绝对值是

$$\begin{cases}|\alpha|=\alpha, & \alpha>0,\\|\alpha|=-\alpha, & \alpha<0,\\|0|=0.\end{cases}$$

实数绝对值的概念是一个重要的概念.从几何上看,它表示该实数到原点的距离,由此引申出欧氏空间中两点间的距离的定义.我们眼前的目的是定义实数的乘法.

13.3.5 实数的积的定义

"积"的定义与"加"的定义实质上相同,但稍微复杂一些.因为要考虑负 × 负 = 正的问题,所以先考虑正实数的乘法.

设已给两个正的实数 α 和 β.我们来考虑一切满足不等式

$$a < \alpha < a', b < \beta < b'$$

的正有理数.

定义 两个正的实数 α 和 β 的积是这样一个实数 γ,它满足不等式

$$ab < \gamma < a'b'. \tag{4}$$

我们证明,这样的数存在而且唯一.

先证存在性.对正有理数,取一切可能的积 ab 的集合.集中的任何数都以任意一个形如 $a'b'$ 的数为上界.令

$$\gamma = \sup\{ab\},$$

则自然有 $ab \leqslant \gamma$,同时还有 $\gamma \leqslant a'b'$.

与和的情形一样,可以用加大数 a, b,缩小数 a', b' 的办法去掉这里的等号,所以 γ 满足积的定义:$ab < \gamma < a'b'$.

接着证明积的唯一性.取正有理数 a, a' 与 b, b',使得,

$$a' - a < e, \ b' - b < e.$$

其中 e 是任意小的正有理数.取数 a', b' 时,要它们分别小于预先给定的数 a_0', b_0':$a' < a_0', b' < b_0'$.于是差数

$$a'b' - ab = a'(b' - b) + b(a' - a) < (a_0' + b_0')e.$$

这就是说,这个差数可以任意地小.由引理 2,只有一个数满足(4).

如果 α 和 β 都是有理数,则它们的普通乘积显然满足(4).

最后,为了定义任意一对实数的积,我们作如下规定.

首先规定，对任意的实数 α，都有

$$\alpha \cdot 0 = 0 \cdot \alpha = 0.$$

如果两个实数都不为 0，则令

$$\alpha \cdot \beta = |\alpha| \cdot |\beta|, \qquad 若\ \alpha, \beta\ 同号;$$

$$\alpha \cdot \beta = -|\alpha| \cdot |\beta|, \qquad 若\ \alpha, \beta\ 异号.$$

和有理数一样，实数保持下面的性质：

1) $\alpha \cdot \beta = \beta \cdot \alpha$;

2) $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$;

3) $\alpha \cdot 1 = \alpha$;

4) $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$;

5) 从 $\alpha > \beta$ 与 $\gamma > 0$ 推得 $\alpha \cdot \gamma > \beta \cdot \gamma$.

13.3.6 实数的商的定义

实数商的定义直接来自实数积的定义.

实数 α 和 β 的商 $\frac{\alpha}{\beta}$ 是这样的数 γ，它满足条件

$$\gamma \cdot \beta = \alpha (或\ \beta \cdot \gamma = \alpha).$$

只要除数 β 不为 0，商就存在且唯一.

这样，我们就定义了实数的四则运算，同时我们还证明了实数系满足域的所有性质，因而实数系构成一个数域.

现在实数的运算就有了严格的逻辑基础. 但是这个严格基础到了 19 世纪才得以完成，是在微积分发现两个世纪之后的事.

古埃及人和巴比伦人早已使用了整数、分数，甚至像 $\sqrt{2}$、$\sqrt{3}$ 这样的无理数；在实际应用中，他们使用无理数的近似值. 但是他们的数学，包括古希腊的数学都是建立在直觉与经验的基础上的，而缺乏严格的逻辑基础.

有趣的是，古希腊人留给后人两门截然不同的、发展得不一样的数学分支. 一门是演绎的、系统的、但有些缺陷的几何学，一门是经验的算术以及它的延伸代数学. 算术和代数却没有自己的逻辑结构. 这

在数学史上形成了一个巨大的反差.

人们在接受整数和小数的性质时,其基础自然是经验.当增添新数时,建立在经验基础上并已为人们接受的整数和小数的运算法则就被应用于这些新数上.算术和代数的理论就这样不断地发展起来,但却没有根基.数学家们为什么没有发展一个数和代数的演绎推理结构呢?这是因为几何的概念、原理和公理在直观上远比代数的概念、定理易于接受.负数、无理数、乃至复数的概念却微妙得多.为数系和代数建立逻辑基础是一个非常困难的问题.这就是数的逻辑基础直到19世纪才建立的原因.

§13.4 根的存在性

13.4.1 具有有理指数的乘幂

由实数乘法和除法的定义可以推出实数的正整数指数与负整数指数的乘幂.事实上,

$$\alpha = \alpha^1, \alpha \cdot \alpha = \alpha^2, \cdots, \alpha \cdot \alpha^n = \alpha^{n+1}.$$

$$\frac{1}{\alpha} = \alpha^{-1}, \alpha^{-1} \cdot \alpha^{-1} = \alpha^{-2}, \cdots, \alpha^{-n} \cdot \alpha^{-1} = \alpha^{-(n+1)}.$$

但是要讨论实数的有理数乘幂,就必需先讨论根的存在问题.

我们还记得,在有理数域中最简单的根都可能不存在,例如$\sqrt{2}$就是一个著名的例子.这是扩充有理数域的一个根据.现在我们已经扩充了有理数域,而引进了实数域.在实数域中可以进行开方运算吗?即对任意实数 α, $\alpha^{\frac{1}{p}} = \sqrt[p]{\alpha}$ 存在吗?我们来证明,它存在.由此,又可定义任意实数 α 的有理数次幂.

设 α 是任一实数,而 n 是自然数.数 α 的 n 次方根是这样的实数 ξ,使得

$$\xi^n = \alpha.$$

我们限定 α 是正数，求满足这个关系式的正数 ξ，即求算术根．我们来证明，这样的数存在且唯一．

唯一性容易证明，因为不同的正数有不同的乘幂：

$$0 < \xi < \xi' \Rightarrow \xi^n < \xi'^n.$$

现在证明存在性．

如果碰巧存在一个正有理数，它的 n 次幂等于 α，则它就是所要求的数 ξ．存在性自然得证．下面我们讨论一般情形．

我们用下面的方法作出全部有理数的分划 $A|A'$．

A 类：一切使 $r^n < \alpha$ 的正有理数 r、零，及一切负有理数．

A' 类：一切使 $r'^n > \alpha$ 的正有理数 r'．

不难看出，分划满足不空、不乱、不漏的要求（请读者验证）．

现在设 ξ 是由分划确定的数．我们要证，$\xi^n = \alpha$，即 $\xi = \sqrt[n]{\alpha}$．

如果 r 和 r' 是有理数，且满足不等式

$$0 < r < \xi < r',$$

即，r 属于 A 类，r' 属于 A' 类，所以

$$r^n < \alpha < r'^n.$$

又，根据实数的不等式性质，

$$r^n < \xi^n < r'^n,$$

我们可以把差数 $r' - r$ 取得小于任意指定的小数 $\varepsilon > 0$．再假定 $r_0' > r'$ 是预先给定的．这时，

$$r'^n - r^n = (r' - r)(r'^{n-1} + r \cdot r'^{n-2} + \cdots + r^{n-1}) < \varepsilon \cdot n r_0'^{n-1}.$$

这就是说，这个数可以任意地小．根据引理 2，ξ^n 与 α 相等．

这就证明了一个正实数可以开 n 次方，其根仍为实数．

证明了根的存在性之后，我们就可以利用通常的方法建立任何有理指数 r 的乘幂的概念了，并且可以验证在初等代数课本中建立的运算法则：

$$\alpha^r \cdot \alpha^s = \alpha^{r+s}, \alpha^r / \alpha^s = \alpha^{r-s}, (\alpha^r)^s = \alpha^{rs}.$$

$$(\alpha\beta)^r = \alpha^r \cdot \beta^r, \left(\frac{\alpha}{\beta}\right)^r = \frac{\alpha^r}{\beta^r}.$$

这里还需指出，当 $\alpha > 1$ 时，α^r 随着有理指数 r 的增大而增大.

13.4.2 任何实指数的乘幂

最后，我们对正实数 α 定义任何实指数 β 的乘幂. 假定 $\alpha > 1$，因为 $\alpha < 1$ 时，可取它的倒数.

对满足不等式

$$b < \beta < b'$$

的有理指数 b，b'，引进 α 的乘幂 α^b. $\alpha^{b'}$. α 的 β 次乘幂，记为 α^β，定义为这样一个数 γ，它满足不等式

$$\alpha^b < \gamma < \alpha^{b'}.$$

用与前面类似的方法可以证明，γ 存在且唯一.

至此魏尔斯特拉斯规划的第一部分已经完成，我们已经为实数系构建了一个严密的逻辑基础. 下面的任务是完成魏尔斯特拉斯规划的第二部分了.

习 题

1. 证明，两个实数的商是存在且唯一的.
2. 证明，$\sqrt{2}\sqrt{3} = \sqrt{3}\sqrt{2} = \sqrt{6}$.

第十四章　极限、连续与积分

彻底回顾一下，我们每次都对过去有不同的看法；我们每次都从过去看出新的方面，我们每次都把重新走过的道路的全部经验加以补充的理解. 充分意识过去，我们就可以认清现在；深深地沉思往事的意义，我们就能发现未来的意义；回顾一下就向前进 ……

赫尔岑

这一讲与实数理论密切相关，是实数理论的延续和加深.

§14.1　极限论

前面已经指出，实数理论是整个微积分的基础，极限、连续、微商和积分的概念都建立在实数理论的基础之上. 现在就按这一顺序建立极限、连续、积分的理论.

极限论要回答的主要问题有两个：一是有哪些判断极限存在的法则，即在什么条件下极限存在；二是如何计算极限，即极限满足哪些基本运算规律. 第二个问题在初等微积分中已讨论过，所以我们将主要研究第一个问题，并且主要讨论序列的极限，因为序列的极限是基础，通过对序列的极限的讨论就可回答上述两个主要问题. 为此，先回顾极限的定义.

定义　称序列 a_n 以 a 为极限，如果对每一个任意小的正数 ε，总存在一个正整数 N，使得对于 $n > N$ 的一切 a_n，不等式

$$|a_n - a| < \varepsilon$$

成立. 常数 a 叫做序列 a_n 当 $n \to \infty$ 时的极限，或者说序列 a_n 收敛到

a,并记作

$$\lim_{n\to\infty} a_n = a.$$

如果序列没有极限,就说序列是发散的.

14.1.1 单调序列

现在我们来研究单调序列.称序列是单调上升的,如果

$$a_1 < a_2 < \cdots < a_n < a_{n+1} < \cdots,$$

也就是说,若 $n < m$ 则 $a_n < a_m$.如果

$$a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n \leqslant a_{n+1} \leqslant \cdots,$$

也就是说,若 $n < m$,则 $a_n \leqslant a_m$,则称序列是不降的(或广义上升的).

类似地,称序列 a_n 是单调下降的,如果

$$a_1 > a_2 > \cdots > a_n > a_{n+1} > \cdots,$$

也就是说,若 $n < m$ 则 $a_n > a_m$.如果

$$a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n \geqslant a_{n+1} \geqslant \cdots,$$

也就是说,若 $n < m$,则 $a_n \geqslant a_m$,则称序列是不升的(或广义下降的).

我们把这种当 n 上升时向一个方向改变的序列称为单调序列.

关于单调序列我们有下面的重要定理.

定理 1 设给定一个单调上升的序列 $\{a_n\}$.如果它有上界:

$$a_n \leqslant M (M = \text{常数}, n = 1,2,3,\cdots),$$

则它必有有限的极限.类似地,给定一个单调下降的序列 $\{a_n\}$.如果它有下界:

$$a_n \geqslant M (M = \text{常数}, n = 1,2,3,\cdots),$$

则它也必有有限的极限.

证 我们只对单调上升(包括广义单调上升)的序列给予证明,因为单调下降的序列的证明是完全类似的.

根据确界定理(见实数理论 §2 定理 3),序列 $\{a_n\}$ 有上确界:

$$a = \sup\{a_n\}$$

我们来证明,数 a 就是变量的极限.事实上,由上确界的性质,对一切 n,

$$a_n \leqslant a,$$

并且,不论取怎样小的正数 ε,总可以找到一个 N,使得,

$$a_N > a - \varepsilon.$$

利用序列的单调性,当 $n > N$ 时,$a_n \geqslant a_N > a - \varepsilon$.所以

$$0 \leqslant a - a_n < \varepsilon \to |a_n - a| < \varepsilon,$$

即,$\lim a_n = a$.证毕.

注 显然地,如果单调上升序列没有上界,则这个序列将趋于无穷.

单调有界变量有极限这一事实,在 19 世纪的前半叶认为是不言而喻的.要求给它以严格证明,是建立无理数的算术理论的一个起源.这个论断与实数连续性的论断是等价的.魏尔斯特拉斯就用这个定理(把当作公理)作为出发点去建立他的实数理论.

这个定理明确给出了极限存在的准则,是一个很有用的定理.数 e 的存在性就是依据这一定理证明的.

例 设 $a_n = \left(1 + \dfrac{1}{n}\right)^n$,证明 $\lim\limits_{n\to\infty} a_n = e$.

解 根据定理 1,我们只需证明:1) 序列 $\{a_n\}$ 是单调的;2) a_n 有界.

1) a_n 的单调性.利用二项式定理,

$$\begin{aligned} a_n &= \left(1 + \frac{1}{n}\right)^n \\ &= 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{1\cdot 2} \cdot \frac{1}{n^2} \\ &\quad + \cdots + \frac{n(n-1)\cdots(n-n+1)}{1\cdot 2\cdots n} \cdot \frac{1}{n^n} \end{aligned}$$

$$= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right)$$
$$+ \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right),$$

如果我们把 a_n 换为 a_{n+1}，那么就要增加一个新项，即第 $(n+2)$ 项，而且每一项的 n 要换为 $n+1$，因此都变大了.所以，

$$a_n < a_{n+1},$$

即 a_n 随 n 的增加是单调上升的.

2) a_n 的有界性.在 a_n 的展式中去掉一切括弧内的因子，就得到

$$a_n < 2 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} = b_n.$$

将此式分母中的每个因子都换成 2，我们得到

$$b_n < 2 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} < 3.$$

从而 $a_n < 3$.有界性得证.利用定理 1 可知，序列有极限.欧拉把这个极限记为 e.e 是数学中最有用的常数之一.

14.1.2 区间套定理

区间套定理是一个很基本的定理，它从另一个角度刻画了实数的完备性.这个定理也十分有用.考虑区间的一个无限序列

$$[a_1, b_1], [a_2, b_2], \cdots, [a_n, b_n], \cdots \tag{1}$$

说(1)构成一个区间套，如果它满足两个条件(图 14－1)：

图 14－1

1) 区间的长度趋于 0，即 $\lim(b_n - a_n) = 0$；

2) 后一个区间总包含在前一个区间内，即 $a_n \leqslant a_{n+1} < b_{n+1} \leqslant b_n$.

定理 2(区间套定理) 若(1)是一个区间套,则存在一个唯一的实数 a 属于一切区间.

证 依假设,我们有,

$$a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n \leqslant \cdots < b_1.$$

从而$\{a_n\}$是一个单调有界序列,根据定理 1,下面的极限存在,设为 a:

$$\lim_{n\to\infty} a_n = a.$$

现在假定 k 是任意一个自然数.若 $n > k$,则

$$a_k \leqslant a_n < b_k.$$

保持 k 固定,让 $n \to \infty$,我们有

$$a_k \leqslant a \leqslant b_k.$$

这就是说 a 属于第k 个区间.由于 k 是任意的,所以 a 属于区间套的一切区间.

数 a 的唯一性是明显的.若我们另有一数 $b > a$,b 也属于区间套的一切区间,这将与区间的长度趋于 0 相矛盾.

注 区间套定理在有理数域内是不成立的.给定一个满足条件的有理数构成的区间套,可能套不住一个有理数,原因是有理数域不完备.

例 设

$$a_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n-1} - \ln n,$$

$$b_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n-1} + \frac{1}{n} - \ln n,$$

证明$[a_n, b_n]$构成一个区间套.

解 只需验证它满足区间套的条件.为此先证一个不等式

$$\frac{1}{n+1} < \ln\frac{n+1}{n} < \frac{1}{n}.$$

事实上，当 $n < x < n+1$ 时，$\frac{1}{n+1} < \frac{1}{x} < \frac{1}{n}$，从而

$$\int_n^{n+1} \frac{1}{n+1} \mathrm{d}x < \int_n^{n+1} \frac{1}{x} \mathrm{d}x < \int_n^{n+1} \frac{1}{n} \mathrm{d}x,$$

积分得

$$\frac{1}{n+1} < \ln \frac{n+1}{n} < \frac{1}{n}.$$

接着证明 a_n, b_n 的单调性.由它们的表达式立刻得到，

$$\begin{aligned} a_{n+1} - a_n &= \frac{1}{n} - \ln(n+1) + \ln n \\ &= \frac{1}{n} - \ln \frac{n+1}{n} > 0, \end{aligned}$$

$$\begin{aligned} b_{n+1} - b_n &= \frac{1}{n+1} - \ln(n+1) + \ln n \\ &= \frac{1}{n+1} - \ln \frac{n+1}{n} < 0. \end{aligned}$$

最后，

$$b_n - a_n = \frac{1}{n} \to 0.$$

这样一来，$[a_n, b_n]$ 构成一个区间套.

根据区间套定理，当 $n \to \infty$ 时，下式

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln n$$

的极限存在.这个极限记为 C，叫做欧拉常数，$C = 0.577\,2\cdots$.这样，我们用区间套定理证明了欧拉常数 C 的存在性.

14.1.3　收敛原理

我们先研究子序列.给定一个序列.

$$a_1, a_2, \cdots, a_n, \cdots, \tag{2}$$

我们可以考虑任何由它的部分项组成的序列，称为它的子序列

$$a_{n_1}, a_{n_2}, \cdots, a_{n_k}, \cdots, \tag{3}$$

其中$\{n_k\}$是某一个上升的自然数的序列：$n_1 < n_2 < \cdots < n_k < n_{k+1} < \cdots$.

不难看出，若序列(2)有极限，则序列(3)也有极限. 但序列(2)没有极限时，序列(3)也可能有极限.

例如，设$a_n = (-1)^n$，则这个变量没有极限. 如果限定n只取偶数，或只取奇数，则子序列分别为

$$a_2 = 1, a_4 = 1, \cdots, a_{2n} = 1, \cdots$$

和

$$a_1 = -1, a_3 = -1, \cdots, a_{2n+1} = -1, \cdots,$$

这两个序列分别以1与-1为它们的极限.

在序列(2)是无界的情况下，要子序列有有限极限有时是不可能的，但是在序列(2)是有界的情况下，一定存在收敛的子序列.

定理3(波尔查诺－魏尔斯特拉斯) 从任何有界的序列(2)中总能选出收敛于有限极限的子序列(3).

证 证明分为两步：1) 用区间套定理找出一个极限点；2) 先取子序列.

1) 设全部数a_n都落在两个有限数A, B之间. 把区间$[A, B]$分为两半，则至少一半包含序列中的无穷多个元素. 如果不然，整个区间将只包含序列中的有限多个元素，这与假定矛盾. 设$[A_1, B_1]$是包含无穷多个a_n的那一半.

同样，从区间$[A_1, B_1]$中又可分出它的一半$[A_2, B_2]$，在它的内部包含无穷多个a_n. 继续这种步骤. 设第k次分出的区间是$[A_k, B_k]$，在它的内部包含无穷多个a_n. 如此继续下去，一直到无穷多次.

从第二个起，这些区间中的每一个都包含在前一个之中，其长度是前一个的一半. 所以，第k个区间的长度是

$$B_k - A_k = \frac{1}{2^k}(B - A),$$

它随着 k 的增加而趋于 0. 由区间套定理,存在实数 c,使得

$$\lim_{k\to\infty} A_k = \lim_{k\to\infty} B_k = c.$$

2) 我们用下面的办法造出子序列 $\{a_{n_k}\}$. 在序列的元素 a_n 中取包含在 $[A_1, B_1]$ 的任一个元素作为 a_{n_1}. 在 a_{n_1} 后面的各元素 a_n 中取包含在区间 $[A_2, B_2]$ 内的任一个元素作为 a_{n_2},如此继续. 一般地说,在以前分出的 $a_{n_1}, a_{n_2}, \cdots, a_{n_{k-1}}$ 的后面各元素 a_n 中取包含在 $[A_k, B_k]$ 内的任一个元素作为 a_{n_k}. 依假设,这种选取是可能的. 这样,我们就得到了一个子序列. 它满足性质

$$A_k \leqslant a_{n_k} \leqslant B_k; \lim_{k\to\infty} A_k = \lim_{k\to\infty} B_k = c.$$

由此,我们立刻得到

$$\lim_{k\to\infty} a_{n_k} = c.$$

这就是所要证明的. 证毕.

这是一个很有用的定理,利用它可以简化某些定理的证明. 下面我们将经常使用它.

我们在前面证明了,单调有界变量一定有极限. 这个定理给出了极限存在的一个判别法. 但是在一般情况下,变量的变化不是单调的,这时也需要知道变量是否有极限. 下面的定理给出了极限存在的充要条件,它具有很大的理论价值. 这个定理属于波尔查诺(1817)与柯西(1821),现在通常把他称为柯西准则.

定理 4 变量 a_n 有极限的充分且必要的条件是:对任意给定的数 $\varepsilon > 0$,存在自然数 N,使得当 $m > N$ 与 $n > N$ 时,下面的不等式总成立:

$$|a_n - a_m| < \varepsilon. \tag{4}$$

证 1) 必要性的证明. 假定变量 a_n 有有限极限 a. 由极限定义,不论 $\varepsilon > 0$ 是怎样的数,对于 $\varepsilon/2$ 可求得自然数 N,使得当 $n > N$ 时,恒有

$$|a_n - a| < \frac{\varepsilon}{2}.$$

现在任意取两个序号 $n > N$ 与 $m > N$,对这两个序号同时有

$$|a_n - a| < \frac{\varepsilon}{2} \text{ 和 } |a_m - a| < \frac{\varepsilon}{2},$$

由此可得,

$$\begin{aligned}|a_n - a_m| &= |(a_n - a) + (a - a_m)|\\ &\leqslant |a_n - a| + |a - a_m| < \varepsilon,\end{aligned}$$

条件的必要性得证.

条件的充分性的证明要困难得多.

2) 充分性的证明. 假定条件成立,并且对于给定的 $\varepsilon > 0$,可以找到这样的序号 N,使得当 $n > N$ 和 $m > N$ 时,不等式(4)成立. 固定 m,我们有

$$|a_n - a_m| < \varepsilon \Leftrightarrow a_m - \varepsilon < a_n < a_m + \varepsilon.$$

从后一个不等式可看出,a_n 是有界的:当 $n > N$ 时,它的数值都包含在数 $a_m - \varepsilon$ 与 $a_m + \varepsilon$ 之间,并且不难放宽这两个界限,使得这前面的 N 个数值 $a_1, a_2, \cdots, a_N$ 也包含在它们的中间.

于是,由定理 3,可以分出一个子序列 $\{a_{n_k}\}$,它收敛到有限极限 c:

$$\lim_{k \to \infty} a_{n_k} = c.$$

我们来证明变量 a_n 也趋于这个极限. 选取大的 k,使得

$$|a_{n_k} - c| < \varepsilon,$$

并使 $n_k > N$. 在(4)中取 $m = n_k$,从而,当 $n > N$ 时,

$$\begin{aligned}|a_n - c| &= |a_n - a_{n_k} + a_{n_k} - c|\\ &\leqslant |a_n - a_{n_k}| + |a_{n_k} - c| < 2\varepsilon,\end{aligned}$$

即
$$|a_n - c| < 2\varepsilon.$$

这就证明了我们的断言.

注 波尔查诺与柯西建立了有限极限存在的充要条件，但是没有给出严格的证明，因为没有实数理论是不能证明它的.

这个定理也叫凝聚原理，因为序列有凝聚之势.

14.1.4 有限覆盖定理

下面我们把有界闭区间$[a,b]$与一个有限的或无限的开区间的集合$\{\sigma\}$放在一起考虑，其中σ表示开区间. 如果对$[a,b]$中的任一点x，都可以找到$\{\sigma\}$中的一个σ，使σ含有x，则称集合$\{\sigma\}$覆盖区间$[a,b]$.

定理5 若有界闭区间$[a,b]$由开区间的无限集合$\{\sigma\}$覆盖，则从$\{\sigma\}$中可以取出有限子集$\{\sigma_1,\sigma_2,\cdots,\sigma_n\}$也覆盖$[a,b]$.

证 反证法. 仍借助对分法. 设区间$[a,b]$不能用有限个开区间σ覆盖. 将区间$[a,b]$分为两半，则至少有一半不存在有限覆盖. 设$[a_1,b_1]$是不存在有限覆盖的那一半. 对$[a_1,b_1]$重复上述过程，得到$[a_2,b_2]$，它也不存在有限覆盖.

无限地继续这个过程，我们就得到一个区间套$[a_n,b_n]$，其中每一个区间都是前一个区间的一半，它们都不存在有限覆盖. 由区间套定理，所有区间$[a_n,b_n]$存在一个公共点c，并且

$$\lim a_n = \lim b_n = c.$$

象$[a,b]$里的其它点一样，点c落在某一个区间σ内. 设这个区间是$\sigma_0=(\alpha,\beta)$，则$\alpha<c<\beta$. 由(5)，当n充分大时，$[a_n,b_n]\subset(\alpha,\beta)$. 这样一来，有一个区间就盖住了$[a_n,b_n]$，故导致矛盾. 证毕.

14.1.5 极限思想的辩证剖析

现在我们稍稍离开严格的证明而对极限思想作些分析. 极限式

$$\lim_{n\to\infty} a_n = a \tag{5}$$

都有哪些含义呢?

1) 有限与无限的相互转化. 从左向右看(5)，是无限向有限的转化. 从右向左看(5)，是有限中包含着无限. 下面是几个例子:

$$1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\cdots=\frac{\pi^2}{6};$$

$$1+\frac{1}{16}+\frac{1}{81}+\frac{1}{256}+\cdots=\frac{\pi^4}{90};$$

$$1+\frac{1}{9}+\frac{1}{25}+\frac{1}{49}+\cdots=\frac{\pi^2}{8}.$$

在学习极限的时候，我们较多地注意到无限向有限转化的这个侧面，而常常忽略有限包含无限这个侧面. 上面的三个式子告诉我们，有限中含有丰富的无限. π 有许多不同的展式就是一个很好的说明.

2) 近似与精确的相互转化. 对任一个具体的 n，(5) 的左边都是右边的一个近似值. 例如

$$1+\frac{1}{9}+\frac{1}{25}(=1.171\ 5\cdots)\approx\frac{\pi^2}{8}(=1.233\ 7\cdots),$$

项数 n 越多，精确度就越高.

定积分是一种和式的极限. 定积分的近似计算就是用有限和去替代极限值，利用了近似与精确这对矛盾的转化.

如果我们从哲学上来看待极限概念，(5) 也有丰富的含义. 首先，它表现了量变质变律：量的变化引起了质的变化. 例如，有理数的序列可以有无理数的极限；正的序列可以有 0 的极限，等等. 还有近似转化为精确，也是量变引起的质变. 其次，它表现了否定之否定律：有限 — 无限 — 有限. 最后，它反映了对立统一律：有限与无限的对立统一；近似与精确的对立统一；质与量的对立统一；运动与静止的对立统一，等等.

极限概念的含义是丰富的，它的多种应用就基于此.

14.1.6 函数的极限

为了研究函数的三个重要性质，我们简单地回顾一下函数极限的定义.

定义　假定函数 $f(x)$ 定义在点 a 的邻域内，若对任意的 $\varepsilon>0$，存在数 $\delta>0$，使得只要 $|x-a|<\delta$，就永远有

$$|f(x)-A|<\varepsilon,$$

就称当 x 趋于 a 时 $f(x)$ 以 a 为极限，记为

$$\lim_{x\to a}f(x)=A.$$

14.1.7　小结

从实数理论到极限，我们证明了下列定理：

1. 戴德金定理：对实数集合的任何分划 $A|A'$，都存在产生这个分划的实数 α；α 或者是上类中的最小数，或者是下类中的最大数.

2. 确界定理：有上界的实数集合，必有唯一的上确界，有下界的实数集合，必有確一的下确界.

3. 单调序列的定理：单调有界序列必有有限极限.

4. 区间套定理：设 $[a_n,b_n]$ 是一个区间套，满足

1) $\lim\limits_{n\to\infty}(b_n-a_n)=0$；　　2) $a_n\leqslant a_{n+1}<b_{n+1}\leqslant b_n$，

则该区间套一定套住一个点.

5. 波尔查诺－魏尔斯特拉斯定理：从任何有界的序列中总能选出收敛于有限极限的子序列.

6. 柯西准则：序列 $\{a_n\}$ 是收敛的，当且仅当 $\forall\varepsilon>0$，$\exists N>0$，使得当 $m,n>N$ 时，永有

$$|a_n-a_m|<\varepsilon.$$

7. 有限覆盖定理：若有界闭区间 $[a,b]$ 由开区间的无限集合 $\{\sigma\}$ 覆盖，则从 $\{\sigma\}$ 中可以取出有限子集 $\{\sigma_1,\sigma_2,\cdots,\sigma_n\}$ 也覆盖 $[a,b]$.

注　上面的 7 个定理构成了整个分析的基础. 这 7 个定理是彼此等价的. 为此，我们作如下证明：确界定理蕴含戴德金定理.

证　设 $A|A'$ 是实数集合的一个分划. 因为 A' 是非空的，所以 A 是有上界的集合. 令

$$\omega = \sup\{A\},$$

则 ω 存在.它是一个实数.由于分划 $A|A'$ 是不漏的,所以 ω 必属于 A 或 A'.若 $\omega \in A$,则它是 A 中的最大数;若 $\omega \in A'$,则它是 A' 中的最小数.

§14.2 函数的连续性

微积分主要是研究连续函数的,因而对连续函数的性质需要有一个清楚的理解和严格的证明.我们将忽略其它细节的讨论,集中力量考察连续函数在闭区间上的三个性质.这就是最大、最小值定理,中间值定理和关于函数一致连续的定理.

14.2.1 中间值定理

我们先给出函数连续的定义.

定义 称函数 $f(x)$ 在点 x_0 是连续的,如果

$$\lim_{x \to x_0} f(x) = f(x_0).$$

这个式子指的是

$$\lim_{x \to x_0} f(x) = f(\lim_{x \to x_0} x),$$

即,极限号可取到函数式内部.

定理 1 设函数 $f(x)$ 定义在闭区间$[a,b]$上,并且是连续的,又在区间的两端点处取异号的数值,则必有一点 c 存在,$a < c < b$,使得

$$f(c) = 0.$$

注 这个定理有简单的几何意义:连续曲线从 x 轴的一侧走到另一侧必与 x 轴相交(图 14 − 2).

证 我们用二分法来证明这一定理.为确定起见,我们假定,$f(a) < 0, f(b) > 0$.

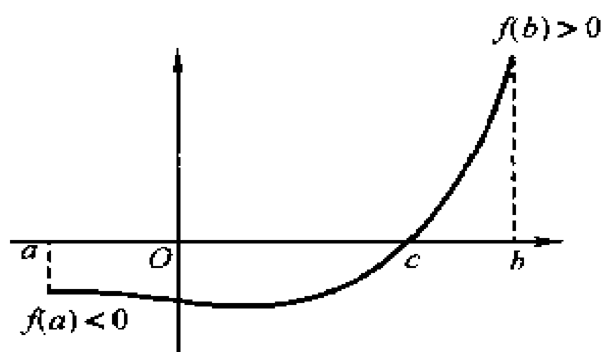


图 14－2

用分点$\frac{a+b}{2}$把$[a,b]$分成两半.若$f(\frac{a+b}{2})=0$,则取$c=\frac{a+b}{2}$,c点已经找到.设$f(\frac{a+b}{2})\neq 0$,则函数必在区间$[a,\frac{a+b}{2}]$或$[\frac{a+b}{2},b]$的两端处取异号的值,并且在左边小于0,在右边大于0.用$[a_1,b_1]$表示这个区间,我们有

$$f(a_1)<0,f(b_1)>0.$$

再用分点$\frac{a_1+b_1}{2}$把$[a_1,b_1]$分成两半.若$f(\frac{a_1+b_1}{2})=0$,则取$c=\frac{a_1+b_1}{2}$,c点已经找到.设$f(\frac{a_1+b_1}{2})\neq 0$,仿上,用$[a_2,b_2]$表示使

$$f(a_2)<0,f(b_2)>0$$

的区间.

继续这种构造的步骤.这时,或者在有限步之后,我们会遇到一个分点,在这一点函数值为0,定理的证明就告完成;或者我们得到一个区间套.

我们来讨论后一种情形.对第n个区间$[a_n,b_n]$,我们有,

$$f(a_n)<0,f(b_n)>0,\text{并且 } b_n-a_n=\frac{b-a}{2^n}.$$

根据区间套定理,存在实数 c,满足

$$\lim a_n = \lim b_n = c.$$

c 显然属于 $[a,b]$. 我们来证明 c 满足定理的要求.

根据函数的连续性,我们有,

$$f(c) = \lim_{n\to\infty} f(a_n) \leqslant 0, f(c) = \lim_{n\to\infty} f(b_n) \geqslant 0.$$

所以,必有 $f(c) = 0$. 证毕.

函数的连续性是一个本质的条件. 若函数不是连续的,则定理不一定成立.

定理 1 在求解代数方程的根上十分有用. 今举一例.

例　证明奇数次实系数代数方程

$$f(x) = x^{2n+1} + a_1 x^{2n} + \cdots + a_{2n} x + a_{2n+1} = 0$$

至少有一个实根.

解　考虑绝对值充分大的 x, 多项式的符号将与 x 的符号相同. 当 $x > 0$ 时,多项式的符号为正;当 $x < 0$ 时,多项式的符号为负. 因为多项式是连续函数,所以它既然变号,必在区间的某一点处取 0 值. 所以奇数次实系数代数方程至少有一个实根.

定理 1 不仅可以用来确定根的存在性,还可以用来计算根的近似值. 前面代数方程式部分已经作过介绍.

将定理 1 稍加推广,我们就得到连续函数的中间值定理.

定理 2　设函数 $f(x)$ 定义在闭区间 $[a,b]$ 上,并且是连续的,又在区间的两端点处取不同的数值

$$f(a) = A, f(b) = B,$$

则不论 C 是 A 与 B 之间的怎样的数,必有一点 c 存在, $a < c < b$, 使得 $f(c) = C$.

证　为确定起见,设 $A < B$, 从而 $A < C < B$. 作辅助函数 $g(x) = f(x) - C$. 这个函数在这区间上是连续的,并且在它的两个端点处取不同的符号:

$$g(a) = f(a) - C = A - C < 0,$$
$$g(b) = f(b) - C = B - C > 0.$$

由定理 1,在 a 与 b 之间可以找到一点 c,使得 $g(c) = 0$,即

$$f(c) - C = 0, \text{或} f(c) = C.$$

证毕.

14.2.2 函数的最大、最小值定理

我们首先注意到,在闭区间上的连续函数都是有界的.

定理 3 若函数 $f(x)$ 定义在闭区间 $[a,b]$ 上,并且是连续的,则它在该区间上是有界的,即存在常数 m 与 M,使得

$$m \leqslant f(x) \leqslant M.$$

证 用反证法.假定函数在区间 $[a,b]$ 上是无界的,不妨设它无上界.

这时,对每一个自然数 n,可在区间 $[a,b]$ 上找到一个数值 x_n,使得

$$f(x_n) \geqslant n. \tag{1}$$

根据 §14.1 定理 3,我们可以从序列 $\{x_n\}$ 中选出一个子序列 $\{x_{n_k}\}$,它收敛到有限极限: $x_{n_k} \to x_0(k \to \infty)$.由于函数在 $[a,b]$ 上是连续的,所以

$$f(x_{n_k}) \to f(x_0).$$

但这是不可能的,因为由(1)式, $f(x_{n_k}) \to \infty$.

这个矛盾就证明了定理.

定理 4 若函数 $f(x)$ 定义在闭区间 $[a,b]$ 上,并且是连续的,则它在该区间上达到它的最大值和最小值.

证 令

$$M = \sup\{f(x)\},$$

根据定理 3,这是一个有限值.假定 $f(x)$ 在区间 $[a,b]$ 上达不到它的上确界 M,那么必有 $f(x) < M$.这时,我们考察辅助函数

$$g(x) = \frac{1}{M - f(x)}.$$

根据假设，分母不为 0，所以这个函数是连续的，因而它是有界的：$g(x) \leqslant N(N > 0)$. 由此可得，

$$\frac{1}{M - f(x)} \leqslant N \Rightarrow M - f(x) \geqslant \frac{1}{N},$$

即

$$f(x) \leqslant M - \frac{1}{N}.$$

这与 M 是 $f(x)$ 的上确界相矛盾. 所得矛盾就证明了定理：在区间 $[a, b]$ 上可以找到 x_0，使得 $f(x_0) = M$.

同样的道理可以证明关于最小值的断言.

注　这又是一个纯粹存在性的证明，在证明过程中没有给出任何求最大、最小值的方法. 我们知道，求最大、最小值的方法是借助微分学来实现的.

求最大、最小值的方法具有广泛的实用价值. 推而广之，就是最优化讨论的课题.

有了定理 4 之后，我们就可以定义函数在一个区间上的振幅了.

若函数 $f(x)$ 在区间$[a, b]$上是有界的，则它在这区间内的上确界与下确界之差

$$\omega = M - m$$

叫做它在这区间内的振幅.

如果函数的值域是$[m, M]$，则振幅就是这个区间的长度.

14.2.3　一致连续性

设函数 $f(x)$ 定义在区间$[a, b]$上，x_0 是区间$[a, b]$内的一点. 若 $f(x)$ 在 x_0 连续，即

$$\lim_{x \to x_0} f(x) = f(x_0).$$

或者用 $\varepsilon - \delta$ 语言说：对于任意给定的 $\varepsilon > 0$，总可以找到这样的数 $\delta > 0$，使得只要 $|x - x_0| < \delta$，就有

$$|f(x) - f(x_0)| < \varepsilon.$$

注意，函数连续性的定义是局部的. 现在假定，函数 $f(x)$ 在区

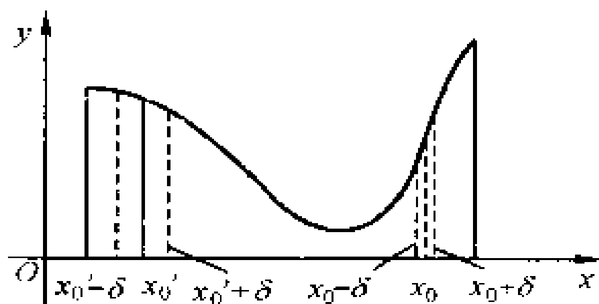


图 14－3

间$[a,b]$的每一点都连续．于是对于区间内的每一点x_0，根据给定的ε可以求得相应的δ．当x_0在区间内变动时，即使ε不变，δ也是要改变的．只要看一看图 14－3 就会明白．在函数变化慢的地方δ就大，在函数变化快的地方δ就小．这就是说，数δ不仅依赖于ε，而且也依赖于x_0．于是出现这样一个问题：当给定ε时，是否存在这样的δ，它适用于这个区间的所有的x_0呢？不一定．这依赖于函数，也依赖于区间的类型．如果存在这样的δ，我们就说，函数在这个区间上是一致连续的；否则就说函数在区间上不是一致连续的．

定义　称函数$f(x)$在一个区间上是一致连续的，若任给$\varepsilon>0$，都能找到一个$\delta>0$，使得对区间内的任意两点x_1,x_2，只要$|x_1-x_2|<\delta$，就有$|f(x_1)-f(x_2)|<\varepsilon$．

那么，一个函数在某个区间上不一致连续如何定义，需要把它弄清楚．许多人在这个问题上常常出错．定理 5 的证明中将使用反面叙述，读者应仔细阅读．

一致连续的概念对数学分析具有重大的意义．连续和一致连续的关系是什么呢？首先，从函数在区间上的一致连续性可以推出它在这个区间上每一点的连续性．这一点毫无困难．困难之点是逆定理：在区间上的连续函数是一致连续的吗？答案依赖于区间的类型．如果区间是开区间，那么一般说来，逆定理不成立．如果区间是闭区间，则

逆定理成立.

定理 5　若函数 $f(x)$ 在区间 $[a,b]$ 上是连续的，则它在这区间内一致连续.

证　我们用反证法证明. 假定对某一个确定的数 $\varepsilon>0$，不存在有一致连续定义中所说的数 $\delta>0$. 这时，不论取怎样的数 $\delta>0$，总可在区间 $[a,b]$ 上找到 x, x'，使得，虽然 $|x-x'|<\delta$，但是

$$|f(x)-f(x')|\geqslant\varepsilon.$$

现在取正数序列 $\{\delta_n\}$，使 $\delta_n\to 0$. 根据所述，对每一个 δ_n 可以在区间 $[a,b]$ 中找到数 x_n, x'_n，虽然 $|x_n-x'_n|<\delta_n$，但是

$$|f(x_n)-f(x'_n)|\geqslant\varepsilon. \tag{2}$$

根据 §14.1 定理 3，从有界序列 $\{x_n\}$ 内可以取出收敛于区间 $[a,b]$ 中某一点 x_0 的子序列. 为简单计，我们认为序列 $\{x_n\}$ 本身就收敛到 x_0.

因为 $x_n-x'_n\to 0$，所以序列 $\{x'_n\}$ 也收敛于 x_0. 由函数在点 x_0 的连续性，

$$f(x_n)\to f(x_0),\ f(x'_n)\to f(x_0),$$

因此，

$$(f(x_n)-f(x'_n))\to 0.$$

但是这与(2) 相矛盾. 证毕.

系　若函数 $f(x)$ 在区间 $[a,b]$ 上是连续的，则对于给定的 $\varepsilon>0$ 可求得这样的 $\delta>0$，只要把区间任意分成长度都小于 δ 的小区间，在小区间内函数的振幅都会小于 ε.

证　对于给定的 $\varepsilon>0$，取在一致连续的定义中所得的那个数作为 δ，则在长度小于 δ 的小区间中，函数的任意两个数值之差按绝对值都小于 ε. 特别地，函数的最大、最小值之差也小于 ε，即函数的振幅都会小于 ε.

这个系将在讨论函数的可积性时用到.

例　证明函数 $f(x)=\frac{1}{x}$ 在开区间(0,1)不是一致连续的.

证明　函数 $f(x)$ 的连续性是明显的，我们来证明它在区间(0,1)上的不一致连续性. 根据定义，我们要证明对某个给定的 $\varepsilon>0$，找不到相应的 δ.

取 $\varepsilon=1$，不论怎样的 $\delta>0$，只要自然数 n 充分大，总可以在(0,1)上找到两点 $x_n=\frac{1}{n}$，$x'_n=\frac{1}{n+1}$，尽管它们的距离小于 δ：

$$|x_n-x'_n|=\frac{1}{n}-\frac{1}{n+1}=\frac{1}{n(n+1)}<\delta,$$

可是

$$|f(x_n)-f(x'_x)|=|n-(n+1)|=1.$$

这就是说，对于 $\varepsilon=1$，我们找不到定义要求的 δ. 证毕.

上例的不一致连续性出在0附近，因为函数在0是无界的.

至此，我们完成了连续函数性质的研究. 有些性质看来是明显的，如连续函数的中间值定理；有的是很精细的，如上面的一致连续性定理. 这些定理的证明只是在实数理论建立之后才有了严格的基础.

§14.3　黎曼积分

我们假定读者已经熟悉初等微积分的性质和计算，因而这里只讨论黎曼积分的存在性问题. 有了前面的准备后，我们可以研究在什么条件下一个函数的定积分存在.

14.3.1　黎曼积分

设函数 $f(x)$ 定义在区间 $[a,b]$ 上. 我们在 a,b 之间插入一些分点，将区间任意分为 n 段：

$$a=x_0<x_1<\cdots<x_n=b.$$

用 λ 表示差数 $\Delta x_i = x_{i+1} - x_i (i = 0,1,2,\cdots,n-1)$ 中最大者.在每个小区间 $[x_i, x_{i+1}]$ 上任取一点 ξ_i,

$$x_i \leqslant \xi_i \leqslant x_{i+1} (i = 0,1,2,\cdots,n-1).$$

作出总和

$$\sigma = \sum_{i=0}^{n-1} f(\xi_i)\Delta x_i.$$

这个和称为积分和.它的值依赖于我们所用的分法,依赖于点 ξ_i 的取法.

现在我们来建立这个有限和的极限概念:

$$I = \lim_{\lambda \to 0} \sigma.$$

定义　如果对于无论怎样小的 $\varepsilon > 0$,都可以找到这样一个 $\delta > 0$,使得对于任意的分法,只要 $\lambda < \delta$,不管怎样选取点 ξ_i,我们都有

$$|\sigma - I| < \varepsilon.$$

那么,我们就称总和 σ 在 $\lambda \to 0$ 时有极限 I. I 叫做函数 $f(x)$ 在区间 $[a,b]$ 上的定积分,记为

$$\int_a^b f(x)\mathrm{d}x. \tag{1}$$

如果这个极限存在,则称函数 $f(x)$ 在区间 $[a,b]$ 上是可积的.

这个一般的定义是黎曼给出的,他最先给出这种一般形式,并研究了它的应用范围,所以和 σ 称为黎曼和.但为了强调它与积分的关系,常把它称为积分和.柯西只是对连续函数研究过积分和.对于不连续函数黎曼作了研究.

现在我们问:在什么条件下,积分和有有限极限,即在什么条件下,积分(1)存在?

我们首先指出,上面的定义只适合于有界函数.事实上,如果一个函数 $f(x)$ 在区间 $[a,b]$ 内是无界的,则在任何分割之下,至少在一个小区间里函数保持无界.于是在这个小区间里总可以选到 ξ,使

$f(\xi)$ 连同总和 σ 大于任何事先指定的数. 在这种情况下, 和 σ 不可能有有限极限. 所以无界函数是黎曼不可积的.

因此, 我们讨论有界函数, 只有它们才可能是黎曼可积的, 并约定

$$a \leqslant x \leqslant b, m \leqslant f(x) \leqslant M.$$

14.3.2 达布和

作为一种辅助工具, 我们引进达布和的概念.

以 m_i 和 M_i 分别表示函数 $f(x)$ 在区间 $[x_i, x_{i+1}]$ 的上确界和下确界, 并做和

$$s = \sum_{i=0}^{n-1} m_i \Delta x_i ; S = \sum_{i=0}^{n-1} M_i \Delta x_i .$$

这两个和各称为下和与上和, 或达布和.

特别地, 当 $f(x)$ 是连续函数时, 这些和对应于任一分法的积分和的最大者和最小者. 因为这时函数 $f(x)$ 在每一个小区间内都能达到它的上确界和下确界, 而点 ξ_i 可以这样取, 使得 $f(\xi_i) = m_i$ 或 $f(\xi_i) = M_i$.

回到一般情形. 由上、下确界的定义, 我们有

$$m_i \leqslant f(\xi_i) \leqslant M_i ,$$

由此可得,

$$s \leqslant \sigma \leqslant S .$$

在固定的分割之下 s 和 S 都是常数, 而和 σ 却是变化的, 因为 ξ_i 可以任意选择, 凭 ξ_i 的选择, $f(\xi_i)$ 的值可以随意接近于 m_i, 也可以随意接近于 M_i, 这就是说, 和 σ 可以随意接近于 s 或 S. 这样一来, 达布和 s 与 S 是积分和 σ 的上确界和下确界.

14.3.3 达布和的性质

达布和有两条简单性质.

性质 1 若在一组现有的分点上增加一些新分点, 则达布上和

有减无增，达布下和有增无减.

证 我们只需看增加一个分点的情形就够了. 设此点落在 x_k 与 x_{k+1} 之间：

$$x_k < x' < x_{k+1}.$$

今用 S' 表示新和. 在和 S 中，对应于 $[x_k, x_{k+1}]$ 的项是

$$M_k(x_{k+1} - x_k).$$

在新和 S' 中对应于该区间的有两项

$$\overline{M}_k(x' - x_k) + \overline{\overline{M}}_k(x_{k+1} - x'),$$

这里 $\overline{M}_k$ 和 $\overline{\overline{M}}_k$ 分别为函数 $f(x)$ 在区间 $[x_k, x']$ 和 $[x', x_{k+1}]$ 内的上确界. 易见，

$$\overline{M}_k \leqslant M_k, \overline{\overline{M}}_k \leqslant M_k.$$

从而

$$\overline{M}_k(x' - x_k) \leqslant M_k(x' - x_k),$$

$$\overline{\overline{M}}_k(x_{k+1} - x') \leqslant M_k(x_{k+1} - x')$$

由此得到，

$$\overline{M}_k(x' - x_k) + \overline{\overline{M}}_k(x_{k+1} - x') \leqslant M_k(x_{k+1} - x_k).$$

由此推出，$S' \leqslant S$. 下和的证法与此类似.

性质 2 每一个达布下和，无论它对应于怎样的分割法，都不大于任一个达布上和.

证 对区间 $[a, b]$ 做任意分割，求出其达布和

$$s_1 \text{ 与 } S_1.$$

对区间 $[a, b]$ 再做另一分割，这一分割与前一分割毫无关系，也求出其达布和

$$s_2 \text{ 与 } S_2.$$

我们要证明，$s_1 \leqslant S_2$. 为此我们把两个分割法的分点并在一起，而得到第三个辅助的分割法，它的相应的达布和为 s_3 与 S_3.

根据性质 1，比较第一分割法与第三分割法，我们有

$$s_1 \leqslant s_3.$$

比较第二分割法与第三分割法，我们有

$$S_3 \leqslant S_2.$$

但 $s_3 \leqslant S_3$，所以我们有，$s_1 \leqslant s_3 \leqslant S_3 \leqslant S_2$，即

$$s_1 \leqslant S_2.$$

证毕.

从上面的证明可以看出，全体下和的集合 $\{s\}$ 是有上界的，例如，任何一个上和 S 就是它的一个上界，所以这个集合有一个上确界

$$I_* = \sup\{s\},$$

称为函数在区间 $[a,b]$ 上的下积分. 并且，对不论哪个上和 S，都有

$$I_* \leqslant S.$$

由此又可知道，上和的集合 $\{S\}$ 有下界. 这样一来，上和的集合 $\{S\}$ 有下确界

$$I^* = \inf\{S\},$$

称为函数在区间 $[a,b]$ 上的上积分. 显然，我们总有下面的关系式

$$s \leqslant I_* \leqslant I^* \leqslant S. \tag{2}$$

14.3.4　积分存在的条件

有了达布和的帮助，我们就容易给出积分存在的条件了.

定理 1　定积分存在的充要条件是

$$\lim_{\lambda \to 0}(S - s) = 0. \tag{3}$$

证　必要性. 假定积分(1)存在，我们要证，对任意给定的 $\varepsilon > 0$，可以找到 $\delta > 0$，使得只要 $\Delta x_i < \delta$，就有

$$|\sigma - I| < \varepsilon \Leftrightarrow I - \varepsilon < \sigma < I + \varepsilon.$$

不论 ξ_i 在相应的子区间上怎样取法. 但是，我们知道，s 与 S 分别是积分和的上、下确界，所以

$$I - \varepsilon \leqslant s \leqslant S \leqslant I + \varepsilon,$$

这就是

$$\lim_{\lambda \to 0} s = \lim_{\lambda \to 0} S = I.$$

从而(3)成立.

充分性.假定(3)成立,由(2),显然有 $I_* = I^*$.以 I 表示它们的公共值,则

$$s \leqslant I \leqslant S.$$

如果把 σ 理解为与 s 及 S 相应于同一分割法的积分和之一,则我们有

$$s \leqslant \sigma \leqslant S.$$

按条件(3),当 $\Delta x_i < \lambda$ 时,$S - s < \varepsilon$.从而,我们有

$$|\sigma - I| < \varepsilon.$$

这就是说,I 是积分和 σ 的极限,即 I 是定积分.证毕.

如果用 $\overline{\omega}_i$ 表示函数在第 i 个子区间的振幅 $M_i - m_i$,则我们有

$$S - s = \sum_{i=0}^{n-1} (M_i - m_i)\Delta x_i = \sum_{i=0}^{n-1} \omega_i \Delta x_i,$$

因而定积分存在的条件(3)可以写成

$$\lim_{\lambda \to 0} \sum_{i=0}^{n-1} \overline{\omega}_i \Delta x_i = 0. \qquad (4)$$

这种形式是一种常用的形式.

14.3.5　可积函数类

应用上面的判别法,我们来确定几类可积分的函数.

1) 若函数 $f(x)$ 在区间 $[a,b]$ 上是连续的,则它是可积的.

证　若函数 $f(x)$ 在区间 $[a,b]$ 上连续,则它在该区间上一致连续.由§14.2 定理 5 的系,对给定的 $\varepsilon > 0$,总可以找到这样的 $\delta > 0$,使得在区间 $[a,b]$ 分为长度 $\Delta x_i < \delta$ 的小区间时,$\overline{\omega}_i < \varepsilon$.从而

$$\sum_{i=0}^{n-1} \overline{\omega}_i \Delta x_i < \varepsilon \sum_{i=0}^{n-1} \Delta x_i = \varepsilon(b - a).$$

注意到,$b - a$ 是常数,而 ε 可以任意小,所以(4)成立,积分存在.

所有的初等函数在它们的定义域内都是连续的，因而在其定义域内的任何闭区间上，初等函数都是可积的.

上面的论断还可以作如下的推广.

2) 若函数 $f(x)$ 在区间 $[a,b]$ 上是有界的，且只有有限个间断点，则它是可积的.

证　我们只需证明函数 $f(x)$ 在区间 $[a,b]$ 上有一个间断点的情形. 设间断点为 x'. 对于任意给定的 $\varepsilon>0$，作邻域 $(x'-\varepsilon,x'+\varepsilon)$（图 14－4）. 从区间 $[a,b]$ 中去掉这个邻域后，剩下两个闭区间，函数在这两个闭区间上一致连续. 对此，我们可以使用 § 14.2 定理 5 的系. 对每一个闭区间都可以找到一个 δ，使得在任一个长度小于 δ 的小区间上，函数 $f(x)$ 的振幅都小于 ε. 根据 ε 所得到的两个 δ 中取较小的一个，仍记为 δ. 我们还可以使 $\delta<\varepsilon$.

现在对区间作分割，使得全部 Δx_i 都小于 δ. 我们把小区间分为两类：

(1) 间断点邻域之外的小区间. 在这种区间里，函数的振幅 $\overline{\omega}_i<\varepsilon$.

(2) 整个小区间或小区间的一部分落在间断点的邻域内.

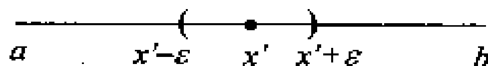


图 14－4

因为函数 $f(x)$ 是有界的，所以函数在任何小区间里的振幅都不会超过它在整个区间 $[a,b]$ 上的振幅 Ω. 我们把总和

$$\sum_i \overline{\omega}_i \Delta x_i$$

分成两个：

$$\sum_1 \overline{\omega}_i \Delta x_i \text{ 和 } \sum_2 \overline{\omega}_i \Delta x_i .$$

分别表示在第一类区间上与第二类区间上的和．对第一个和，我们有

$$\sum_1 \overline{\omega_i}\Delta x_i < \varepsilon \sum \Delta x_i < \varepsilon(b-a).$$

至于第二个和，我们注意到，完全落在间断点邻域内的小区间的总长不超过 2ε；落在间断点邻域的端点处的小区间的个数最多有两个，它们的总长度是 $2\delta < 2\varepsilon$ 所以，

$$\sum_2 \overline{\omega_i}\Delta x_i < \Omega \sum \Delta x_i < \Omega \cdot 4\varepsilon.$$

这样一来，我们有

$$\sum_i \overline{\omega_i}\Delta x_i < \varepsilon[(b-a)+4\Omega].$$

注意到方括弧里的是常数，ε 可任意小，命题得证．

3）单调函数 $f(x)$ 是可积分的．

证　我们只需考虑单调增函数，单调减函数的证明是完全一样的．设 $f(x)$ 单调增函数．于是它在区间 $[x_i, x_{i+1}]$ 上的振幅是

$$\omega_i = f(x_{i+1}) - f(x_i).$$

对任意给定的 $\varepsilon > 0$，取

$$\delta = \frac{\varepsilon}{f(b)-f(a)}.$$

当 $\Delta x_i < \delta$ 时，我们有

$$\sum \omega_i \Delta x_i < \delta \sum [f(x_{i+1}) - f(x_i)] = \delta[f(b)-f(a)] = \varepsilon.$$

这就证明了单调增函数是可积的．

注　可积函数在有限个点上的函数值的变化不影响积分的存在性，也不影响积分值．这是因为，这种变化只涉及到总和中的有限项．当 $\lambda \to 0$ 时，该和仍趋于 0．至于积分的值，我们在选取 ξ_i 时，总可避开那些函数值．

有了定积分的存在定理、定积分的理论就建立在牢固的基础上了．有没有不可积的函数呢？如果没有不可积的函数，那么我们就没有必要研究可积函数类了．事实上，存在不可积的函数．今举一例．

例　考虑狄利克雷函数 $D(x)$,其定义如下:

$$D(x)=\begin{cases}1, & x\in R\\ 0, & x\in I\end{cases}$$

$x\in R$ 表示 x 是有理数,$x\in I$ 表示 x 是无理数.显然,$D(x)$ 在区间 $[0,1]$ 上是有界的,但它却不可积.因为,无论把区间分得多么细,在每个小区间 $[x_i,x_{i+1}]$ 中总有有理数,也总有无理数,于是,$M_i=1$,$m_i=0$,$\overline{\omega_i}=1$.从而总有

$$S-s=1.$$

由定理 1,$D(x)$ 不可积.

一个函数是黎曼可积的充要条件是勒贝格给出的.这要用到测度的概念,也就是需要对点集论作精细论述.这超出了本书的范围.定理的叙述是这样的:

定理(勒贝格)　函数 $f(x)$ 在区间 $[a,b]$ 上是可积的充分且必要的条件是,它是有界的,并且它在 $[a,b]$ 上的不连续点所成的集合的测度为 0.

粗糙地说,如果一个函数的间断点,不管有多少(可能很多),都可用一些小区间盖住,这些小区间的总长度可以任意小,那么这个函数就是可积的.

现在可以说,微积分已经有了严密的逻辑基础.

第十五章　数学模型

一种科学只有在成功地运用数学时，才算达到完善的地步.

——马克思

甚至一个粗糙的数学模型也能帮助我们更好地理解一个实际的情况，因为我们在试图建立数学模型时被迫考虑了各种逻辑可能性，不含混地定义了所有的概念、并且区分了重要的和次要的因素. 一个数学模型即使导出了与事实不符合的结果，它也还可能是有价值的，因为一个模型的失败可以帮助我们去寻找更好的模型.

——A. Renyi

数学模型属于应用数学，它涉及到纯数学与其它学科的交互作用. 本讲将提供几个典型的颇具启发性的实例，以说明数学在社会的各个领域中的应用情况. 这一章冠以"数学模型"一词，实在是因为它已成为应用数学的一大分支，而目前正处于蓬勃发展的时期. 它的本义就是将各种各样的实际问题化为数学问题.

解决实际问题的步骤分为以下五个阶段：

1）科学地识别与剖析实际问题；

2）形成数学模型；

3）求解数学问题；

4）研究算法，并尽量使用计算机；

5）回到实际中去，解释结果.

数学家在第一阶段起不到明显的作用，起作用的通常是研究这类问题的科学家、工程师、医生、甚至是企业家. 正是这些人认识到了问题的重要性和与数学方法的可结合性. 由于近年来数学的应用已引起广泛的注意，所以常常是这样，在提出系统的理论以前，有关数

据的收集，经验性的结论已完成．所欠的是数学家的介入，数学家的介入将会使问题发生质的变化．

第二阶段是整个建模过程中最困难最关键的部分．它最富有创造性，常由具有数学知识的科学家参加，或由数学家与科学家共同参与．模型的建立由仔细地理解问题，区分主次和选取合适的数学结构所组成．模型有两个方面，一是数学结构，一是实际概念与数学结构间的对应．在建模过程中，必须保留原始问题的本质特征，但要尽可能地简化．注意，简化是基于科学而不是基于数学．简化是必须的，以便使得到的数学体系是容易处理的．但又不能过分，以防数学定理不能提供实际情形的有效预测．决定什么是重要的，什么是不重要的；哪些简化是合理的，哪些简化是不合理的，需要经验与技巧．需要科学家与数学家共同来完成．

基于对同一问题的观察和研究，提出的数学模型可能有几种不同数学结构．不同数学结构可能反映问题的不同侧面．例如，光的物理模型有两个，一个是波动说，一个是粒子说．它们都是有用的．

第三个阶段是求解数学问题．这个阶段的研究在表面上与纯数学的研究没有区别，只是动机不同．但是，这里的数学问题与实际问题有密切的联系，记住这一点很重要．一但所提问题由于数学自身的原因需要修改时，必须仔细分析修改后的问题与实际问题之间的关系．

看来简单的问题引出来的数学问题未必简单，有可能引出极难的数学问题．常常是这样，实际问题的研究为数学打开了一个全新的领域，导致创立新的数学分支．有时某些问题能自然地融合进我们熟悉的数学课题中，这自然很令人愉快．我们下面讨论的模型都属于这种情况．读者切勿错误地理解为数学模型都是这样．恰恰相反，更多的情形不是这样．

第四阶段的计算是另一个重要的阶段．为了获得对原问题的理解，计算的结果是不可少的．由于实际问题的复杂性，大部分结果是不能借助手工来完成的，所以算法的研究以及使用计算机是必须的．

最后的阶段是依照原问题去解释和评价所得结果．这时可能出现各种情况，我们需要作仔细分析．这就推动我们去进一步完善模型．

下面就来具体介绍几个典型的数学模型，它们分别来自政治学、考古学、人口学、运动员训练，以及经济学等不同领域．

§15.1 选票分配

选举问题是政治学研究的中心问题之一．其中包括民意测验，选票分配等重要问题．我们从著名的选举悖论谈起，为此，先简单谈谈什么叫悖论．

15.1.1 何谓悖论

‘悖论’这个词的含义比较丰富，它包括一切与人的直觉和日常经验相矛盾的结论．悖论有三种主要形式：

1）一种论断看起来好像是错了，但实际上却是对的．这是一种似非而是的论断(佯谬)．

2）一种论断看起来好像是对的，但实际上却是错的．这是一种似是而非的论断．

3）导致逻辑上自相矛盾的论断．

悖论具有重要的哲学意义和数学意义．从古希腊的芝诺提出的悖论开始，一直到罗素的关于集合论的悖论，都对数学理论的发展起了巨大的推动作用．

15.1.2 选举悖论

假定有张、王、李三个同学竞选学生会主席．民意测验表明，两两比较，选举人中有 2/3 愿意选张不愿选王，有 2/3 愿意选王不愿选李．问：关于张和李我们应该得出什么结论呢？是不是愿意选张而不愿选李的人多呢？

答案是：不一定！如果选举人按照表 15－1 那样对候选人进行排

序，就会引起一个惊人的悖论.

表 15 - 1

	1	2	3
$\frac{1}{3}$	张	王	李
$\frac{1}{3}$	王	李	张
$\frac{1}{3}$	李	张	王

现在我们对他们进行两两的比较.

张和王的民意测验情况是：张有两次排在王的前面，而王只有一次排在张的前面，因而张可以说，选举人中有 2/3 人喜欢我.

王和李的民意测验情况是：王有两次排在李的前面，因而王可以说，选举人中有 2/3 人喜欢我.

李和张的民意测验情况是：李有两次在张的前面，而张只有一次在李的前面，因而李也可以说，选举人中有 2/3 人喜欢我.

这就出现了一个令人惊讶的悖论：多数人选举人选张优于王，多数选举人选王优于李，还是多数选举人选李优于张.

在日常生活中，许多关系都是可以传递的. 例如，大小关系. $A > B$，$B > C$，就可以推出 $A > C$. 还有上下关系，前后关系，左右关系等等. 所有这些关系都具有传递性. 这就使人们以为"好恶"关系也是可以传递的. 但实事上，"好恶"关系是不可以传递的.

这个悖论可追溯到十八世纪，它是一个非传递关系的典型，这种关系在人们作两两对比选择时可能产生.

这条悖论有时称作阿洛悖论. 肯尼思·阿洛曾根据这条悖论和其他逻辑理由证明了，一个十全十美的民主选举系统是不可能实现的. 这就是说，不存在公平合理的选举系统. 这是一个非常深刻的结论，但更加有悖于常理：天下竟然无公！这个结论告诉我们，只有更公，没有最公. 阿洛因此分享了 1972 年的诺贝尔经济学奖.

假定有三个对象,而且具有三种可以比较的指数,将它们按各指标排好顺序,当我们进行两两比较时,就可能出现上述矛盾.假定张、王、李是向一个姑娘求婚的三个人,表15-1所示的排列可解释为这个姑娘就三个方面比较这三个人的优劣次序,例如第一是学位,第二是容貌,第三是收入.如果两两比较,这个可怜的姑娘就会发现她觉得张比王好,王比李好,李又比张好!

这个悖论还可以在产品检验中出现.一个统计学家也许会发现,2/3年青妇女喜欢润肤霜 A 超过 B,2/3年青妇女喜欢润肤霜 B 超过 C.化学公司得知这一结果后也许就将润肤霜 C 作为最不受欢迎的一种而降低产量.殊不知,第三个统计可能会表明还有2/3的人喜欢 C 超过 A 呢.

15.1.3 选票分配问题

选票分配问题属于民主政治的范畴.选票分配是否合理是选民最关心的热点问题之一.这一问题早就引起西方政治家与科学家的关注,并进行了大量深入的研究.这项研究大量地使用了数学方法.本节以学生会选举为例,对这项研究作一初步介绍.

我们知道,每个高等院校都有学生会,学生会大约每四年改选一次.在每届学生会改选时,都需要给出各系的委员分配名额.那么名额怎么分配才算合理呢?按照学生会章程规定,各系的委员数按学生人数比例分配.

假定某大学的学生会由 n 名委员组成,再设该大学有 s 个系,各系的学生数是 p_i,$i=1,2,\cdots,s$.全校的学生数是

$$p=p_1+p_2+\cdots+p_s.$$

现在的问题是,找出一组相应的整数 $n_1,n_2,\cdots,n_s$,使得

$$n_1+n_2+\cdots+n_s=n,$$

其中 n_i 是第 i 个系获得的委员数.

按照学生会章程,一个简单而公平的分配委员名额的办法是按

人数比例分配. 记

$$q_i = \frac{p_i}{p} \cdot n,$$

称它为分配的份额. 自然有

$$q_1 + q_2 + \cdots + q_n = n.$$

如果 q_i 都是整数,分配不会出现问题. 但是更经常发生的情况是,q_i 不是整数,而名额分配又必须是整数,怎么办?一个自然想到的办法是"四舍五入法". 四舍五入的结果可能会出现名额多余,或名额不够的情况. 我们举例来说明这一情况. 假定某学院有三个系,总人数是200,下面三个表说明了三种不同情况:按四舍五入法,表 15 - 2 正好产生 20 个委员;表 15 - 3 产生 19 个委员;表 15 - 4 产生 21 个委员.

表 15 - 2

系别	学生数	所占比例(%)	按比例分配名额	最终分配名额
甲	107	53.5	10.7	11
乙	59	29.5	5.9	6
丙	34	17	3.4	3
总和	200	100	20	20

表 15 - 3

系别	学生数	所占比例(%)	按比例分配名额	最终分配名额
甲	104	52	10.4	10
乙	62	31	6.2	6
丙	34	17	3.4	3
总和	200	100	20	19

表 15 - 4

系别	学生数	所占比例(%)	按比例分配名额	最终分配名额
甲	105	52.5	10.5	11
乙	60	30	6	6
丙	35	17.5	3.5	4
总和	200	100	20	21

三张表表明三种情况.一般说来,用四舍五入法很难得到所需要的委员数.这说明四舍五入法有缺陷,需要改进,以找出更合理的分配法.

15.1.4 亚拉巴马悖论

正因为四舍五入法有这一缺点,美国乔治·华盛顿时代的财政部长亚历山大·汉密尔顿于 1790 年提出了一种解决名额分配的办法,并于 1792 年为美国国会所通过.美国国会的议员是按州分配.汉密尔顿方法的具体操作如下:

1) 取各州份额的整数部分$[q_i]$,先让第 i 州拥$[q_i]$有个议员.

2) 然后考虑各个 q_i 的小数部 $q_i-[q_i]$,按从大到小的顺序将余下的名额分配给相应的州,直到名额分配完为止.

下表是按照汉密尔顿方法进行分配的:

表 15 - 5

系别	学生数	所占比例(%)	按比例分配名额	最终分配名额
甲	103	51.5	10.3	10
乙	63	31.5	6.3	6
丙	34	17	3.4	4
总和	200	100	20	20

汉密尔顿方法看起来十分合理,但是仍存在问题.例如在表 15 - 4 中出现了甲和丙的小数部分相同的情况.如果甲系和丙系各增加一个名额就会出现超出预定名额的情况;如果两系都不增加,就会出现名额不满的情况.当然在选民众多的情况下,出现小数部分相等的情况是十分罕见的,换言之,概率很小.但是下面提到的亚拉巴马悖论,却是必须严肃对待的情况.

从 1880 年起,美国国会就汉密尔顿方法的公正合理性展开了争论.原因是 1880 年美国人口普查后,亚拉巴马州发现汉密尔顿方法触犯了该州的利益,其后 1890 年和 1900 年人口普查后,缅因州和科罗拉多州也极力反对汉密尔顿方法.按照常规,假如各州的人口比例

不变,议员的名额总数由于某种原因而增加的话,那么各州的议员名额数或者不变,或者增加,至少不应该减少.可是汉密尔顿方法却不能满足这一常规.这里还以校学生会为例.由于考虑到20个名额的成员在表决提案时可能会出现10:10的结果,所以学生会决定下届增加一个名额.按照汉密尔顿方法分配名额得到表15-6.

表15-6

系别	学生数	所占比例(%)	按比例分配名额	最终分配名额
甲	103	51.5	10.815	11
乙	63	31.5	6.615	7
丙	34	17	3.570	3
总和	200	100	21	21

计算结果表明,总名额增加一个,丙系反而减少一个名额,这当然侵犯了丙系的利益.亚拉巴马州当年就面临这种情况,所以通常把汉密尔顿方法所产生的这一矛盾叫做亚拉巴马悖论.

这个悖论是出乎意料的,它是在实践过程中出现的,不是逻辑的产物.

因此,必须进一步改进汉密尔顿方法,使之更加合理,新的方法不久就提出来了,并消除了亚拉巴马悖论.因方法较繁杂,这里不再作详细介绍.但是新方法又引出了新问题,新问题又需要消除,于是更新的方法,当然是更加公正合理的方法又出现了,这更新的方法也还有问题.于是出现了这样的问题:是否有一种公正合理的分配方法呢?

这个问题从诞生之日起,就一直吸引着众多政治家与数学家去研究.这里要特别提出的是巴林斯基和扬两位学者,他们在名额分配的研究中引进了公理化方法,并于1982年证明了关于名额分配的一个不可能定理,即包括"不产生人口悖论","不违反'公平分配'原则"等在内的五条十分合理的公理不相容.换言之,满足这五条公理的名额分配方法是不存在的.

§15.2 体育训练问题

用现代数学方法研究体育运动是从 20 世纪 70 年代开始的. 1973 年,美国的应用数学家 J.B.开勒发表了赛跑的理论,并用他的理论训练中长跑运动员,取得了很好的成绩. 几乎同时,美国的计算专家艾斯特运用数学、力学,并藉助计算机研究了当时铁饼投掷世界冠军的投掷技术,从而提出了他自己的研究理论,据此提出了改正投掷技术的训练措施,从而使这位世界冠军在短期内将成绩提高了 4 米,在一次奥运会的比赛中创造了连破三次世界纪录的辉煌成绩. 这些例子说明,数学在体育训练中也在发挥着越来越明显的作用. 所用到的数学内容也相当深入,这里不可能作详细介绍. 我们选择一个较为简单的例子作一说明.

在铅球投掷的训练中,教练关心的核心问题是投掷距离. 如所周知,距离的远近主要取决于两个因素:速度和角度. 这两个因素中哪个更重要呢?为此,我们建立一个数学模型来讨论这一问题.

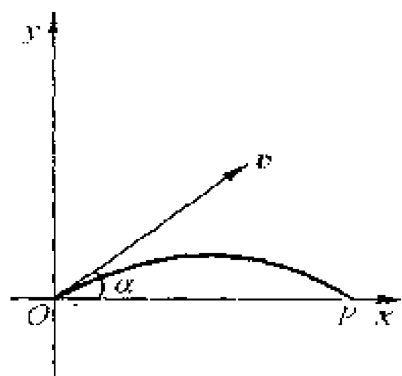


图 15-1

我们在这一模型中不考虑铅球运动员在投掷区域内身体的转动,只考虑铅球的出手速度与投射角度. 此外,还给出三条假定:

1) 忽略铅球在运行过程中的空气阻力作用;

2) 投射角度与投射初速度是相互独立的;

3) 将铅球视为一个质点.

先考虑铅球从地平面以初速度 v 和角度 α 投掷出,如图 15-1 所示,铅球在 P 点处落地. 现在我们来求铅球的运动方程. 设铅球在时

刻 t 的动点坐标为 $(x(t), y(t))$. 我们得到运动方程：

$$\begin{cases} x = v\cos\alpha \cdot t, \\ y = v\sin\alpha \cdot t - \frac{1}{2}gt^2. \end{cases}$$

方程中含有参变量 t，消去 t 才能得到关于 x、y 的关系式. 为此，将代 $t = x/(v\cos\alpha)$ 入 y 中，得

$$y = -\frac{g}{2v^2\cos^2\alpha}x^2 + \tan\alpha \cdot x.$$

这就是运动方程. 为了求出铅球落地处的坐标，令 $y = 0$，解得

$$x_1 = 0,$$

$$x_2 = \frac{2v^2\sin\alpha\cos\alpha}{g} = \frac{v^2\sin 2\alpha}{g},$$

其中 x_1 是铅球起点的 x 坐标，x_2 是铅球落地点 P 的 x 坐标. 若 v 固定，则投掷距离是投射角 α 的函数. 当 $\alpha = 45°$ 时，$\sin 2\alpha = 1$，投掷距离达到其最大值，这时的投掷距离为 $\frac{v^2}{g}$. 这就是说，按 45° 角投掷时，投掷的距离最远.

但实际上铅球不是从地面上出手，而是从一定的高度 h 处出手. 因而上面的运动方程应调整为

$$\begin{cases} x = v\cos\alpha \cdot t, \\ y = v\sin\alpha \cdot t - \frac{1}{2}gt^2 + h. \end{cases}$$

将 $t = x/(v\cos\alpha)$ 代入 y 中，得到

$$y = -\frac{g}{2v^2} \cdot \frac{x^2}{\cos^2\alpha} + \tan\alpha \cdot x + h.$$

令 $y = 0$，得方程

$$\frac{g}{2v^2\cos^2\alpha} \cdot x^2 - \tan\alpha \cdot x - h = 0,$$

解得

$$x_{1,2}=\frac{\tan\alpha\pm\sqrt{\tan^2\alpha+4hg/(2v^2\cos^2\alpha)}}{\frac{g}{v^2\cos^2\alpha}}$$

$$=\frac{v^2\sin 2\alpha}{2g}\pm\sqrt{\left(\frac{v^2}{2g}\sin 2\alpha\right)^2+h\cdot\frac{2v^2}{g}\cos^2\alpha}.$$

舍去负根,我们得到点 p 的 x 坐标为

$$x=\frac{v^2\sin 2\alpha}{2g}+\sqrt{\left(\frac{v^2}{2g}\sin 2\alpha\right)^2+h\cdot\frac{2v^2}{g}\cos^2\alpha}.$$

利用这一公式,列表给出速度与角度对投掷距离的影响(见表15－7).

表 15－7

速度 v(米/秒)	角度 α(度)	距离 x(米)
11.5	47.5	14.929
11.5	45	15.103
11.5	42.5	15.182
11.5	40	15.169
11.5	38	15.092
11.5	36	14.96
11.5	41.2	15.187
11.5	41.6	15.189
11	41.6	14.032
12	41.6	16.395

从表15—7中可以看出,当 $v=11.5$ 米/秒时,最佳角度为41.6°.当角度 α 在38°到45°间变化时,产生的距离差是0.097米.角度的16%的偏差引起距离的0.06%偏差.速度从11米/秒变到12米/秒引起了距离从14.032米/秒到16.395米/秒的偏差,这就是说,速度9%的增加导致了距离16.8%增加.这个结果表明,教练在训练运动员时,应集中主要精力来增加投掷的初始速度.

上面的模型比较粗糙，还有许多问题没有考虑到，例如运动员的身体转动问题，投掷者的手臂长度，肌肉的爆发力，铅球的质量等，还有其它的因素．加上以上诸因素后，得出的公式自然会更精确，当然也会复杂得多．这里不再作深入介绍．铁饼与标枪的投掷问题，篮球投篮问题也属于类似的模型，可用类似上面的方法进行研究．目前在国外，运用数学方法研究体育项目的训练问题已经深入到体育运动的各个领域．这很值得我们借鉴．

§15.3 指数增长与衰减问题

15.3.1 一个简单的微分方程

在实际生活中有许多量，它随时间的变化率正比于它的大小．例如，银行的存款按一定的利率增加；世界的人口按照一定的增长率增加．当然还有别的例子，下面将陆续介绍．

在数学上恰有一个函数能描述上述现象，这就是指数函数．指数函数关于自变量的变化率正比于它的大小：

$$\text{若}, y = Ce^{kx}, \text{则} \frac{dy}{dx} = ky(k \text{ 是常数}). \tag{1}$$

因此，用指数函数来描述上述现象我们将不会惊讶．事实上，满足方程(1)的函数一定是指数函数．

定理 若$\frac{dy}{dx} = ky$，则 $y = Ce^{kx}$，这里 C 是任意常数．

证 由(1)，$\frac{y'}{y} = k$，

从而
$$\int \frac{y'}{y} dx = \int k dx,$$
$$\ln y = kx + C_1,$$
$$y = e^{kx+C_1} = e^{C_1}e^{kx} = Ce^{kx}(C = e^{C_1}).$$

定理得证.

这样一来,我们证明了

$$\frac{dy}{dx} = ky \Leftrightarrow y = Ce^{kx}.$$

我们刚才解的方程(1)是一个含有函数的导数的方程,人们称这种方程为微分方程.微分方程的解是函数,而不是数,这是与代数方程不同的地方.

例 一种细菌按这样的方式增加:在每个时刻它按小时计算的增长率等于它现有总量的两倍.问,一个小时后这种细菌的总量是多少.

解 细菌的增长是离散的,而不是连续的.因为它的总量很大,可以当作连续量来处理仍有很高的近似度.设细菌在时刻 t 的总量为 $y(t)$.依题设它满足微分方程

$$\frac{dy}{dt} = 2y.$$

再设细菌的初始总量是 y_0.因此解为 $y = Ce^{2t}$,但 $y(0) = y_0$,所以

$$y_0 = Ce^0 = C.$$

从而解为 $y = y_0e^{2t}$.当 $t = 1$ 时,

$$y = y_0e^2 \approx 7.4y_0.$$

这样一来,一小时后,细菌的总量将是原有量的 7.4 倍.

15.3.2 人口模型

人口问题是一个极关重要的问题.它强烈地影响着一个地区、一个民族或一个国家的经济发展,现在已引起了世界各国的普遍关注.我国是世界上人口最多的国家,对人口问题也极为关注.研究人口增长的规律,对制定经济政策十分重要.我国的人口政策是,提高人口质量,控制人口数量.

用数学来描述人口增长的设想至少可以追溯到 18 世纪.近代人口问题研究的先驱是英国的经济学家马尔萨斯(Malthus,T.R 1766

—1834).1798 年他发表了《人口论》,引起了广泛的注意.他认为,人口增长的趋势永远快于生产的增长.如果不加限制,人口总是按几何级数增长,而生产资料总是按算术级数增长.当人口扩张到生产资料仅能维持生存的极限时,就会出现饥馑、战争和疾病.马尔萨斯的理论主要有两个错误,一是按照他的模型,人口的数量将趋于无穷;二是人口的增长将引起饥馑、战争和疾病.因此,他的理论受到了批评和修正.五十年代末,我国著名的经济学家马寅初先生提出了《新人口论》,引起了广泛的注意,对我国人口政策的制定有重要影响.

我们从马尔萨斯的模型谈起.马尔萨斯对人口的增加作了两条假定:1) 人口数量是连续变化的量;2) 人口增长的瞬时速度与人口当时的数量成正比.尽管人口的增加或减少是离散的,但在人口量很大的情况下,作为连续量来处理仍能很好地符合于客观情况.这样我们可以假定人口是时间的连续函数,甚至它是 t 的可微函数.由此,我们就可以得到人口方程.

设 $y(t)$ 表示 t 时刻某地区的人口数,则 $y'(t)$ 就表示人口数随时间的变化率.用 $k(t)$ 表示出生率和死亡率的差.如果该地区的人口孤立的,即没有移进移出的移民,则人口的变化率 $y'(t)$ 等于 $ky(t)$.在大多数情况下,可以假定 k 是常数,即不随时间变化.这样一来,我们又得到方程

$$\frac{\mathrm{d}y}{\mathrm{d}t} = ky.$$

这一方程在人口学中叫做马尔萨斯定律.前面已指出,其解为

$$y = Ce^{kt},c \text{ 为任意常数}.$$

如果在 t_0 时,某地区的人口数为 y_0,则

$$y_0 = Ce^{kt_0}, C = y_0 e^{-kt_0},$$

代入前一方程,得

$$y(t) = y_0 e^{k(t-t_0)}. \tag{2}$$

这个解明确而简单,但需要考察一下它是否符合实际情况.我们来看

看全世界人口增长的情况.

从 1960 年到 1970 年世界人口的平均年增长率是 2% . 我们从这十年的中间一年,1965 年 1 月算起. 根据美国财政部的估计,这时全世界的人口总数是 33.4 亿. 因而 $t_0 = 1965$, $y_0 = 33.4$ 亿, $k = 0.02$, 于是

$$y(t) = 33.4 \times 10^8 e^{0.02(t-1965)}.$$

检查这一公式的办法之一是,计算世界人口翻一翻所需要的时间,并与观察值 35 年做一比较.

根据公式(2),若 T 年后地球的人口翻一翻,则

$$2y_0 = y_0 e^{0.02T},$$
$$e^{0.02T} = 2,$$
$$0.02T = \ln 2,$$
$$T = 50\ln 2 \approx 34.6(\text{年})$$

这个值与过去的观察值是十分相近的.

尽管如此,我们还希望用这个公式预见一下更远的未来. 根据公式,到 2515 年世界人口将是 2 000 000 亿,到 2625 年将是 18 000 000 亿,到 2660 年将是 36 000 000 亿. 这些天文数字的意义很难理解. 整个地球面大约 80% 为水所覆盖. 假定我们也愿意在船上生活,那么到 2515 年每个人将仅有 0.87 平方米;到 2625 年每个人将仅有 0.09 平方米;到 2660 年,每个人肩上还得站两个人.

看来这个模型还有不合理处,似乎应把它抛弃掉. 但是,且慢. 因为这个公式与过去的事实是非常相符的,所以我们不能轻率地就将它抛弃. 而且,我们看到了,许多实例都说明人口确实按指数增长. 因而我们的任务是修改模型. 这将在下面论述.

15.3.3 再论人口模型

在人口数量不很大的时候,方程(1) 还是很精确地反映了人口增长的实际情况. 但是当人口数量变得很大时,这一方程的精确程度

就降低了.因为这时人口数量将受到环境因素的很大影响.这些环境因素包括自然资源、食物、居住条件以及战争、瘟疫等,特别地,也包括人口的自我控制.这样一来,我们的方程里应该有一项反映这一环境因素.统计结果告诉我们,在方程中应当加一项 $-by^2$,这里 $b>0$ 是一个常数.因此,我们考虑修改后的方程

$$\frac{\mathrm{d}y}{\mathrm{d}t}=ky-by^2. \tag{3}$$

这个方程叫人口增长率方程.k,b 叫做生命系数.它是 1837 年首先由荷兰的数学—生物学家弗尔哈斯特(Verhulst)引进的.常数 b 相对于k 而言是一个很小的数,所以当 y 不很大的时候,$-by^2$ 这一项与 ky 相比可以忽略.但是当 y 很大时,$-by^2$ 这一项就不容忽略了,它降低了人口增长的速度.不必说,工业化程度越高的国家,生存空间就越大,食物就越丰富,常数 b 就越小.

现在我们应用这一方程去预测人口的增长.设 t_0 时的人口数量是 y_0,时刻 t 的人口数量是$y(t)$.于是我们有初值问题

$$\begin{cases}\dfrac{\mathrm{d}y}{\mathrm{d}t}=ky-by^2,\\ y(t_0)=y_0.\end{cases} \tag{4}$$

现在我们面临两上任务:1)解方程(4);2)用得出的解去预测未来的人口发展.为了求出它的解,将方程变形为

$$\frac{\mathrm{d}y}{ky-by^2}=\mathrm{d}t.$$

两边对 t 积分,可得

$$\int_{y_0}^{y}\frac{\mathrm{d}y}{kt-by^2}=\int_{t_0}^{t}\mathrm{d}t=t-t_0. \tag{5}$$

为了求出左边的积分,需要把被积函数分解为更简单的函数.为此,设

$$\frac{1}{ky-by^2}=\frac{1}{y(k-by)}=\frac{A}{y}+\frac{B}{k-by}, \tag{6}$$

其中 A,B 是待定常数,只要求出 A,B 的值,分解式就得到了.上式两边乘 y,得

$$\frac{1}{k-by}=A+\frac{By}{k-by}.$$

令 $y=0$,得出 $A=\frac{1}{k}$.再用 $(k-by)$ 乘(6)两边,得

$$\frac{1}{y}=\frac{A}{y}(k-by)+B.$$

令 $y=k/b$,可得 $B=b/k$.这样 A,B 都已定出.由此我们有

$$\begin{aligned}\int_{y_0}^{y}\frac{\mathrm{d}y}{y(k-by)}&=\frac{1}{k}\int_{y_0}^{y}\frac{1}{y}\mathrm{d}y+\frac{b}{k}\int_{y_0}^{y}\frac{1}{k-by}\mathrm{d}y\\&=\frac{1}{k}\ln y\Big|_{y_0}^{y}-\frac{1}{k}\ln|k-by|\Big|_{y_0}^{y}\\&=\frac{1}{k}\ln\frac{y}{y_0}+\frac{1}{k}\ln\left|\frac{k-by_0}{k-by}\right|\\&=\frac{1}{k}\ln\left|\frac{y}{y_0}\cdot\frac{k-by_0}{k-by}\right|.\end{aligned}$$

回到(5),我们得出

$$k(t-t_0)=\ln\left|\frac{y}{y_0}\cdot\frac{k-by_0}{k-by}\right|.$$

当 $t>t_0$ 时,上式左边为有限正数,所以 $k-by_0\neq 0$.由此可知当 $t_0<t<+\infty$ 时,$k-by(t)$ 无零点(分母不会为 0).这样一来 $k-by(t)$ 不会改变符号.即

$$\frac{k-by_0}{k-by}>0.$$

因此

$$k(t-t_0)=\ln\frac{y}{y_0}\cdot\frac{k-by_0}{k-by}.$$

两边取指数,可得

$$e^{k(t-t_0)} = \frac{y(k - by_0)}{y_0(k - by)}.$$

或 $$y_0(k - by)e^{k(t-t_0)} = y(k - by_0).$$

把含 y 的项移在左边，我们有

$$[k - by_0 + by_0 e^{k(t-t_0)}]y(t) = ky_0 e^{k(t-t_0)}.$$

因此

$$y(t) = \frac{ky_0}{by_0 + (k - by_0)e^{-k(t-t_0)}}. \tag{7}$$

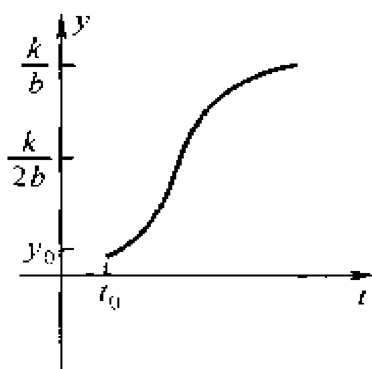


图 15－2

现在我们对(7)作一考察，看看它对人口发展作了一种什么样的预测.

先看长期效果.当 $t \to \infty$ 时，会出现什么情况呢?由(7)可推得，

$$y(t) \to \frac{k}{b}.$$

这样一来，不管人口的初始值是什么，它总是趋向于一个极限值 $\frac{k}{b}$. 曲线 $y(t)$ 的形状如图 15－2 所示.这条曲线叫做 S 形曲线或人口数量增长曲线.这条曲线以 $y = \frac{k}{b}$ 为水平渐近线.

其次，从图 15－2 我们还可以看出下面的事实:人口的数量始终

是增加的;开始增加得慢,然后逐渐加快,在 $y=\frac{k}{2b}$ 附近增加得最快,以后又减缓下来,趋于一个极限.

现在我们对上面的几何结论给以分析的证明.

函数 $y(t)$ 的递增性是明显的,因为 $e^{-k(t-t_0)}$ 是 t 的单调减函数.

再者,方程

$$\frac{dy}{dt}=ky-by^2$$

的两边对 t 求导,得

$$\frac{d^2y}{dt^2}=k\frac{dy}{dt}-2by\frac{dy}{dt}=(k-2by)\frac{dy}{dt}=(k-2by)(k-by)y.$$

由此可见,

$$y<\frac{k}{2b}\Rightarrow\frac{d^2y}{dt^2}>0\Rightarrow\frac{dy}{dt}\uparrow;$$

$$y=\frac{k}{2b}\Rightarrow\frac{d^2y}{dt^2}=0;$$

$$y>\frac{k}{2b}\Rightarrow\frac{d^2y}{dt^2}<0\Rightarrow\frac{dy}{dt}\downarrow.$$

于是,$y=\frac{k}{2b}$ 是一个转折点.从图形我们可以得到这样的结论,在人口达到其极限值的一半之前,是加速增长时期.过了这一点,人口的增长率就减低,这是减速增长的时期,人口的增长率最后趋于0.

为了利用我们的结果去预测地球上未来的人口数量,我们必须估计方程中的生命系数 k 和 b.某些生态学家估计 k 的自然值是0.029.我们已经知道,当人口总数为 3.34×10^9 时,人口的增长率是2%.人口的增长率是指 y'/y.由

$$\frac{1}{y}\frac{dy}{dt}=k-by,$$

我们有

$$0.02=k-b(3.34)\times10^9=0.029-3.34b\cdot10^9$$

因此，$b=2.695\times10^{-12}$. 这样一来，地球上的总人数将趋于极限值

$$\frac{k}{b}=\frac{0.029}{2.695\times10^{-12}}=107.6\text{ 亿}.$$

根据这一预测，世界人口仍处在加速增长时期.

1998 我们曾利用公式(7) 预测地球上到 2000 年将会有多少人口. 令 $k=0.029, b=2.695\times10^{-12}, y_0=3.34\times10^9, t_0=1965, t=2000$. 我们有

$$y(2000)=\frac{0.029\times3.34\times10^9}{0.009+0.02e^{0.029\times35}}$$
$$=\frac{29\times3.34}{9+20e^{-1.015}}\times10^9=59.6\text{ 亿}.$$

15.3.4 三论人口模型

在人口模型中，参数的选取也是一个很重要的问题. 要通过各种实例反复检验才能确定它们. 上面的 k, b 的选取看来还需要改进. 因为按上面的取法，2000 年世界人口总量将是 59.6 亿. 这个估计保守了. 那时估计到 1999 年中期世界人口总量将达到 60 亿. 实际上，1999 年 10 月 12 日被宣布为“世界 60 亿人口日”.

对人口增长作近期估计，更好的办法是用最小二乘法找经验公式. 为此，先简单地介绍最小二乘法. 在实际工作中常常需要根据测量所得的一组数据来找出函数关系，即经验公式. 最小二乘法就是一种寻找经验公式的方法. 我们先通过一个简单实例来说明它.

例 已知金属棒的长度 l 与温度 t 有关. 若用 l_0 表示 0℃ 时金属棒的长度，则金属棒的长度 l 随温度 t 的变化规律是

$$l=l_0(1+ct) \tag{8}$$

这里 c 是金属的膨胀系数.

公式(8) 称为理论公式. 要确定它只需要定出常数 l_0 与 c. 今测得金属棒的长度 l 与温度 t 之间有如下五组数据

温度 t(℃)	20	30	40	50	60
长度(毫米)	1000.36	1000.53	1000.74	1000.91	1001.06

试用这五组数据找出与 c 的最佳值.

解　令 $a = l_0, b = l_0 c$,则(8)可写为

$$l = a + bt \tag{9}$$

因为测量总有误差,因此把上面的数据代入(9)时,得到五个方程,而只有两个未知数,一般不会有准确解.因而我们把目标放在找最佳近似解上.

把 $t_i(i = 1,2,3,4,5)$ 的值代人(9),得到五个理论值 $a + bt_i$,它们和观测值 l_i 都有误差,误差为 $|l_i - a - bt_i|$.我们要找一组最佳的值使总误差最小,即各个误差的和最小.但是绝对值不便作微分运算,所以我们将绝对值改为平方.这种根据各个误差的平方和为最小的条件来确定系数的方法,称为最小二乘法.

于是问题化为求函数 $S = \sum_{i=1}^{5}(l_i - a - bt_i)^2$ 的最小值点 (a,b).由极值的必要条件 $\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0$ 得

$$-2\sum(l_i - a - bt_i) = 0,$$

$$-2\sum(l_i - a - bt_i)t_i = 0.$$

整理后得

$$5a + \left(\sum_{i=1}^{5} t_i\right)b = \sum_{i=1}^{5} l_i,$$

$$\left(\sum_{i=1}^{5} t_i\right)a + \left(\sum_{i=1}^{5} t_i^2\right)b = \sum_{i}^{5} l_i t_i. \tag{10}$$

把有关数据代人方程计算出来,得到

$$5a + 200b = 5003.6,$$

$$200a + 9\,000b = 200\,161.8.$$

解得 $a = 1\,000.01, b = 0.001\,78$. 由此得到公式

$$l = 1\,000.01 + 0.017\,8t,$$

并可算出金属棒的膨胀系数 $c = 0.000\,017\,8$.

上面是利用最小二乘法求直线型经验公式的例子.有时数据间的关系不是一次函数,而是二次函数,幂函数或指数函数.在这种情况下仍可以用最小二乘法求经验公式.

现在回到主题,仍考虑人口问题.我们已经知道,人口的增长规律是指数型函数,所以可以设

$$y(t) = e^{a+bt} \tag{11}$$

根据 n 组数据 $(t_i, y_i)(i = 1,2,\cdots,n)$ 可以用最小二乘法求出 a 和 b 的最佳值.事实上,我们可以在公式(11)两边取对数,得

$$\ln y = a + bt$$

这时数据 (t_i, y_i) 应换为 $(t_i, \ln y_i)$.然后借助最小二乘法求出 a, b 的最佳值.

例 据统计,20 世纪 60 年代世界人口增长情况如下表:

年	1960	1961	1962	1963	1964	1965	1966	1967	1968
人口亿	29.72	30.61	31.51	32.13	32.34	32.85	33.56	34.20	34.83

试求出最佳拟合曲线,并预测公元 2000 年时的世界人口.

解 利用最小二乘法,我们求出 a, b 的最佳值是,

$$a = -26.4258, b = 0.01757$$

最后,我们得到

$$y = e^{26.4258+0.01757t}.$$

2000 年世界人口总数的预测值是 60.8876 亿.

这个结果可能比实际人口多一点.显然,用 80 年代的人口数据作预测,结果会更准确.

在人口学中还有许多其它问题需要研究,例如,妇女人口的增长规律,老年人口的分布情况等等,这些问题对于制定一个国家的经济

计划都有重要的影响.

习　题

1. 设瑞士人口以每年 0.2% 的连续增长率按指数函数规律增长,并设 1988 年的人口总量为 6.6 百万.试写出人口作为时间(以年为单位)的函数表达式.

2. 细菌的增长率与它的总数成正比.若细菌总数在 24 小时内由 100 万增长为 400 万.问,前 12 小时后的总数是多少?

3. 流行病的传染速度问题.考虑这种情况:一个人一旦被感染,与他直接接触的人也立刻被感染.如果人口作正常的流动,那么我们可以假定,流行病传染的速率正比于被感染的人口数(他们是传播细菌者),同时也正比于未被感染的人口数.

如果被感染的人数在全体人口总数中所占的比例由 $\frac{1}{2}$ 增加到 $\frac{2}{3}$ 需要一个月的时间.试问再过一个月,病人占多大比例?

15.3.5　新产品销售模型

一种新产品面世,厂家和商家总要采取各种措施,包括大做广告等,促进销售.他们都希望对产品的销售速度与销售数量做到心中有数,以便用于组织生产,安排进货.

我们的目的是安排一个数学模型来描述产品推销速度,并由此分析出有用结果,以指导生产与销售.

我们讨论耐用商品,这种商品可以长期使用,价格较高,一般不会废弃和重复购置,价格一般也相对稳定.这一类型的新产品,例如微波炉、电饭煲等,刚进入市场时,人们对其功能尚不熟悉,所以销售速度较慢.随着销售数量的增加,人们对于它的熟悉程度就会增加,销售速度也增加,但当这类商品销售到一定数量时,因为人们不会重

复购置,而使销售速度减慢.假设需求量有一个上界 M,用 $x(t)$ 表示时间 t 已售出的产品数量,则尚未购置的数量大约为 $M-x(t)$. 社会调查表明,产品的销售速度$\frac{dx}{dt}$与销售量$x(t)$和 $M-x(t)$ 的乘积成正比,若比例系数记为 k,则

$$\frac{dx}{dt}=kx(M-x). \tag{12}$$

这个方程实质上与方程(3)一样,只是系数稍有不同.用同样的方法可解得

$$x(t)=\frac{M}{1+Ce^{-kMt}} \tag{13}$$

其中 C 是任意常数.它的图形与图 15-2 一样.

对(12)求导,得

$$\frac{d^2x}{dt^2}=k\frac{dx}{dt}(M-x)-kx\frac{dx}{dt} \tag{14}$$

$$=k\frac{dx}{dt}(M-2x).$$

当 $x=M/2$ 时,

$$\frac{d^2x}{dt^2}=0.$$

从(13),可求出 t_0,使 $x(t_0)=M/2$.

由此可做如下分析:

1) 当 $t<t_0$ 时,$x''(t)>0$,因此 $x'(t)$ 单调上升;

2) 当 $t>t_0$ 时,$x''(t)<0$,因此 $x'(t)$ 单调下降.

这样一来,$x'(t)$ 在 $t=t_0$ 时达到最大值.这表明,在销售量小于最大销售量的一半时,销售速度是不断增大的,销售量达到最大销售量的一半时,产品最为畅销,其后销售速度开始下降.

15.3.6 牛顿冷却定律

现在我们研究一个来自热学的问题:物体冷却的定律.牛顿经过

长期的实验和观察得到下面的定律.

牛顿冷却定律 物体的冷却率正比于物体温度与房间温度的差.

与物体相比,若房间非常大时,可假定房间的温度保持不变,即保持常温(这等价于假定物体不是火炉一样的东西,它不会改变房间的温度),也就是物体对房间温度的改变可以忽略不计.

假定房间的温度是 A,物体在 t 时刻的温度是 $y(t)$. 再设 $y_0 = y(0)$. 依牛顿冷却定律,我们有下述形式的微分方程:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = k(A - y), \tag{15}$$

再令 $B = a - y_0$,则方程(15)满足初始条件的解为

$$y = A - Be^{-kt}. \tag{16}$$

把(16)代入(15)直接验证就知道它是解. 我们还是推导一下. 为此引进一个新的变量 $u = A - y$,于是

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{\mathrm{d}y}{\mathrm{d}t} = -k(A - y) = -ku,$$

即新变量 u 满足方程

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -ku,$$

其初始条件为 $u(0) = A - y_0 = B$,因此,解为

$$u = Be^{-kt}.$$

从而

$$y = A - u = A - Be^{-kt}.$$

在使用公式(16)时,常将它化成下述形式:

$$kt = \ln\frac{A - y_0}{A - y} = -\ln\frac{y - A}{y_0 - A}. \tag{17}$$

例 用开水去泡速溶咖啡,3 分钟后咖啡的温度是 85℃. 如果房间的温度是 20℃,问多少分钟后,咖啡的温度降到 60℃?(忽略杯子的冷却影响.)

解　我们有 $A = 20, y_0 = 100$，以及

$$\frac{dy}{dt} = k(20 - y),$$

这里 y 是时刻 t 的咖啡的温度．因为

$$y_0 - A = 100 - 20 = 80,$$

由(17)，

$$kt = -\ln\frac{y-20}{80}.$$

当 $t = 3$ 时，$y = 85$．这时 $y - 20 = 65$．因此

$$k = -\frac{1}{t}\ln\frac{y-20}{80} = -\frac{1}{3}\ln\frac{65}{80} = \frac{1}{3}\ln\frac{16}{13}.$$

当 $y = 60$ 时，$y - 20 = 40$．这时

$$t = -\frac{1}{k}\ln\frac{y-20}{80} = -\frac{3}{\ln\frac{16}{13}}\ln\frac{40}{80}$$

$$= \frac{3}{\ln\frac{16}{13}}\ln 2 \approx 10(分).$$

咖啡从 100℃ 降到 85℃ 花了 3 分钟，降到 60℃ 花 10 分钟．因而从 85℃ 降到 60℃ 需花 7 分钟．

§ 15.4　在考古学中的应用

15.4.1　放射性年龄测定法

测定考古发掘物年龄的近代方法与本世纪初发现的放射现象密切相关．物理学家卢瑟福(Rutheford) 和他的同事们证明了，一些“放射性”元素的原子是不稳定的，在给定的一段时间内，有一个固定比例的原子自然蜕变而形成新元素的原子．卢瑟福证明了，一种物质的蜕变率正比于该物质现有的原子数．因此，如果用 $N(t)$ 表示 t 时刻

的原子数，则$\frac{\mathrm{d}N}{\mathrm{d}t}$表示单位时间内原子的蜕变数，并且它与 N 成正比，即

$$\frac{\mathrm{d}N}{\mathrm{d}t}=-\lambda N, \tag{1}$$

这里 λ 是一个正常数，称为该物质的衰变常数. 自然，λ 越大，物质蜕变的越快. 衡量一种物质蜕变率的一个尺度是它的半衰期. 半衰期定义为给定数量的放射性原子蜕变一半所需要的时间.

现在我们借助 λ 来计算物质的半衰期. 为此，设在 $t=t_0$ 时 $N(t_0)=N_0$. 于是，问题化为求

$$\frac{\mathrm{d}N}{\mathrm{d}t}=-\lambda N, N(t_0)=N_0 \tag{2}$$

的解. 初值问题(2)的解是

$$N=N_0\mathrm{e}^{-\lambda(t-t_0)}. \tag{3}$$

读者一定注意到了，这一问题的解与人口问题的解是一样的. (3)等价于

$$\frac{N}{N_0}=\mathrm{e}^{-\lambda(t-t_0)},$$

两边取对数，可得

$$-\lambda(t-t_0)=\ln\frac{N}{N_0}\Rightarrow t-t_0=\frac{1}{\lambda}\ln\frac{N_0}{N}. \tag{4}$$

利用(4)我们可以很容易地找出半衰期与 λ 的关系. 如果用 T 表示某种放射性物质的半衰期，在(4)中令 $N=\frac{1}{2}N_0$，则

$$T=\frac{1}{\lambda}\ln 2 \text{ 或 } \lambda=\frac{1}{T}\ln 2.$$

这样一来，物质的半衰期是 ln 2 除以衰变常数 λ，λ 的量纲是时间的倒数. 例如，碳 − 14 的半衰期是 5568 年，铀 − 238 的半衰期是 4.5×10^8 年. 对碳 − 14 而言，

$$\lambda = \frac{1}{5586}\ln 2.$$

“放射性测定年龄法”的主要根据是方程(4).如果 t_0 是物质最初形成或制造出来的时间,则物质现在的年龄是 $\frac{1}{\lambda}\ln\frac{N_0}{N}+t_0$.在大多数情况下,衰变常数是已知的,而且 N 的值是容易算出的.因此知道了,就能确定物质的年龄.下面我们来介绍一种年龄测定法.

测定考古发掘物年龄的最精确的方法之一是大约在1949年W·利贝(Libby)发明的碳-14(^{14}C)年龄测定法.这个方法的依据令人愉快地简单.地球周围的大气层不断受到宇宙射线的轰击.这些宇宙射线使地球中的大气产生中子,这些中子同氮发生作用而产生^{14}C.因为^{14}C会发生放射性衰变,所以通常称这种碳为放射性碳.这种放射性碳又结合到二氧化碳中在大气中漂动而被植物吸收.动物通过吃植物又把放射性碳带人它们的组织中.在活的组织中,^{14}C的摄取率正好与^{14}C的衰变率相平衡.但是,当组织死亡以后,它就停止摄取^{14}C,因此^{14}C的浓度因^{14}C的衰变而减少.地球的大气被宇宙射线轰击的速率始终不变,这是一个基本的物理假定.这就意味着,在像木炭这样的样品中,^{14}C原来的蜕变速率同现在测量出来的蜕变速率是一样的(不得不提到的是,20世纪50年代中期以后,核武器试验使得大气中放射性碳的数量显著增加了).有了这个假设我们就能够测定木碳样品的年龄了.设 $N(t)$ 表示在时刻 t 样品中存在的^{14}C的数量,N_0 表示在时刻 $t=0$ 时样品中的数量,即样品形成时的数量.若 λ 是^{14}C的衰变常数(^{14}C的半衰期是5568年),则由(3),我们有

$$N(t) = N_0 e^{-\lambda t}.$$

由此得到样品^{14}C中目前的蜕变率 $N'(t)$:

$$N'(t) = -\lambda N(t) = -\lambda N_0 e^{-\lambda t}$$

而原来的蜕变率是 $N'(0) = -\lambda N_0$.因此

$$\frac{N'(t)}{N'(0)} = e^{-\lambda t}.$$

从而
$$t = \frac{1}{\lambda}\ln\frac{N'(0)}{N'(t)} = \frac{T}{\ln 2}\ln\frac{N'(0)}{N'(t)}. \tag{5}$$

所以如果我们测出木炭中^{14}C目前的蜕变率 $N'(t)$，并注意到 $N'(0)$ 必须等于相当数量的活的树木中^{14}C的蜕变率，那末，我们就能算出木炭的年龄.

例 试确定马王堆一号墓的年代.

解 长沙马王堆一号墓于1972年8月出土，当时曾引起国内外的轰动.现在我们使用碳－14年代测定法来测定它的年代.

开墓时测得木炭中碳－14的平均原子蜕变数是29.78次／分.新木炭的平均原子蜕变数是38.37次／分.把这些数据代入(5)就可以求出该墓的大致年代.事实上，$N'(0) = 38.37$，$N'(t) = 29.78$，$T = 5568$，从而

$$t = \frac{5568}{\ln 2}\ln\frac{38.37}{29.78} = 2036(\text{年}).$$

这样就估计出马王堆一号墓的大致年代是2000年前(西汉末年).

15.4.2 范·米格伦伪造名画案[①]

二次世界大战期间，在比利时解放以后，荷兰保安部开始搜捕纳粹同谋犯.在曾把大量艺术品卖给德国人的某商号的档案中，他们发现了一个银行家的名字，这个银行家曾充当把17世纪荷兰名画家杨·弗美尔(Jan Vermeer，1632—1675)的油画"捉奸"卖给戈林的中间人.这个银行家又泄露，他是第三流荷兰画家H·A·范·米格伦(Van Meegeren)的代表，因此范·米格伦因通敌罪于1945年5月29日被捕.同年7月范·米格伦在牢房里宣布，他从未把"捉奸"卖给戈林，并说，这幅画和非常著名，非常美丽的"埃牟斯的门徒"，以及其它四幅冒充弗美尔的油画和两幅冒充德胡斯(de Hooghs，17世纪荷兰

① 本节内容引自 M.Braun 的 Dfferential Equations and their Applications 一书(Springer-Verlag，1993，第4版)——作者

画家）的油画都是他自己的作品. 这件事震惊了全世界. 但是许多人都认为范·米格伦不过是在撒谎，以免被定为叛国罪. 范·米格伦为了证实他所说的话，他在监狱里开始伪造弗美尔的油画"耶稣在医生们中间"，向怀疑者们证实，他是伪造弗美尔作品的高手. 当这项工作几乎要完成的时候，范·米格伦获悉，通敌罪已改为伪造罪. 因此，他拒绝最后完成这幅油画，并使它变陈，以使满怀希望的检查者们不能发现他使伪造品变陈的秘密. 为了澄清这一问题，由一些卓越的化学家、物理学家和艺术史学家组成的国际专门小组受命调查这一事件. 他们用 X 射线检查画布上是否曾经有过别的画. 此外，他们分析了颜料，考查了画中有没有经历岁月的痕迹.

不过，范·米格伦是很晓得这些方法的. 为了避免别人发觉，他从不很值钱的古画上刮去颜料而只用画布，然后设法用弗美尔使用过的颜料. 范·米格伦也知道，陈年颜料是很坚硬的，而且不可能溶解. 因此他很机灵地在颜料里掺了一种叫酚醛类人工树脂的化学药品，这在油画完成后在炉子上烘乾时就硬化为酚醛树脂.

但是，范·米格伦的伪造工作有几点疏忽之处，使专家小组找到了现代颜料钴兰的痕迹. 此外，他们在几幅画里检验出 20 世纪初才发明的酚醛类人工树脂. 根据这些证据，范·米格伦于 1947 年 10 月 12 日被确认为伪造罪，判刑一年. 服刑期间他因一次心脏病发作而死于 1947 年 12 月 30 日.

但是，即使知道了专家组收集的证据之后，许多人还是不肯相信"埃牟斯的门徒"是范·米格伦伪造的. 他们的论据是，其它所谓的伪造品以及范·米格伦最近完成的"耶稣在医生们中间"质量都是很低的，他们肯定，美丽的"埃牟斯门徒"的作者不会画出质量如此之低的作品来. 事实上，"埃牟斯的门徒"曾被著名的艺术史学家 A·布雷丢斯(Bredius) 鉴定为弗美尔的真迹，并且被伦布兰特(Rembrandt)学会以 170 000 美元的高价购去. 专家小组对怀疑者的答复是，由于范·米格伦曾因他在艺术界没有地位而十分沮丧，他决心绘出"埃牟

斯的门徒”以证明他高于第三流画家.当创作出这样一幅杰作之后，他的志气消退了.而且，当他看到“埃牟斯的门徒”多么容易卖掉以后，在泡制后来的伪制品时就不太用心了.这种解释不能使怀疑者们感到满意.他们要求一个完全科学的，判定性的证明，指出“埃牟斯的门徒”的确是伪制品.卡内基·米伦大学的科学家们在 1967 年做到了这一点.现在我们来叙述他们的工作.

测定油画依旧需要放射性的知识.我们从中学化学所熟知的事实说起，地壳中几乎所有岩石都含少量的铀.岩石中的铀蜕变为另一种放射性元素，而该放射性元素又蜕变为一系列其它元素，最后变为无放射性的铅.铀的半衰期是 4.5×10^9 年，它不断为这一系列中后面的各元素提供来源，使得当它们蜕变后就有前面的元素予以补充.

所有的油画中都含有少量的放射性元素铅 - 210(^{210}Pb)，还有更少量的镭 - 226(^{226}Rd)，因为两千多年来画家用的颜料铅白，即氧化铅都含有这种元素.铅白是从金属铅冶制成的，而金属铅又是从铅矿石中提炼出来的.在提炼过程中，矿石中的铅 - 210 随同金属铅一起被提炼出.但是，90% ～ 95% 的镭以及它蜕变的后裔则随同其它废料作为矿渣而被除去.这样，铅 - 210 的绝大部分来源被切断，它便以 22 年的半衰期非常迅速地蜕变.这个过程一直进行到铅白中的铅 - 210 同所余少量的镭再度处于放射性平衡为止，这时铅 - 210 的蜕变恰好被镭的蜕变所补足而得到平衡.

现在让我们利用这一信息，根据制造铅白时的原有铅 - 210 的含量来计算铅 - 210 现在在样品中的含量.设 $y(t)$ 是在时刻 t 每克铅白所含铅 - 210 的数量.y_0 是制造时间 t_0 的每克铅白所含铅 - 210 的数量，$r(t)$ 是时刻 t 每克铅白中的镭 - 226 每分钟蜕变的数量.如果 λ 是铅 - 210 的衰变常数，则我们有下面的方程：

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y + r(t), y(t_0) = y_0 \tag{6}$$

因为我们关心的时间至多 300 年,而镭 − 226 的半衰期是 1600 年,我们可以假定镭 − 226 保持不变,所以 $r(t)$ 是一个常数 r. 这样(6)变为

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y + r, y(t_0) = y_0. \tag{7}$$

我们现在来求解这一初值问题. 先求方程的一般解. 我们有

$$\frac{\mathrm{d}y}{-\lambda y + r} = \mathrm{d}t$$

两边积分:

$$\int \frac{\mathrm{d}y}{-\lambda y + r} = \int \mathrm{d}t,$$

求解得:

$$-\frac{1}{\lambda}\ln|\lambda y - r| = t + C_1,$$

$$\ln|\lambda y - r| = -\lambda t - \lambda C_1,$$

$$|\lambda y - r| = e^{-\lambda C_1} \cdot e^{-\lambda t}.$$

令 $C = \pm e^{-\lambda C_1}$,则

$$\lambda y - r = Ce^{-\lambda t},$$

$$y = \frac{C}{\lambda}e^{-\lambda t} + \frac{r}{\lambda}. \tag{8}$$

这就得到了方程的一般解. 为了求出特解,利用条件 $y(t_0) = y_0$ 来定出常数 C:

$$y_0 = \frac{C}{\lambda}e^{-\lambda t_0} + \frac{r}{\lambda},$$

$$C = (\lambda y_0 - r)e^{\lambda t_0}.$$

代入(8),得出特解

$$y = \frac{r}{\lambda}(1 - e^{-\lambda(t-t_0)}) + y_0 e^{-\lambda(t-t_0)}. \tag{9}$$

现在就从这个公式出发来判别油画的真伪. $y(t)$ 和 r 可以很容易地测量出来. 因此,如果我们知道了 y_0,就能利用(9)来计算 $t - t_0$,由此我们就可确定油画的年龄. 但是,我们不能直接测量 y_0,克服这一

困难的方法之一是利用下述事实：在用来提炼金属铅的矿石中，铅 - 210 的原始含量与较多的镭 - 226 处于放射平衡．因此，让我们取不同矿石的样品，并计算各矿石中每分钟镭 - 226 蜕变的原子数．对不同矿石所做的结果列在表 15 - 9 中．这些数在 0.18 到 140 之间变化．因此，在刚生产出来时，每克铅白中所含铅 - 210 每分钟蜕变的原子数在 0.18 到 140 之间变化．这意味着 y_0 也在一个很大的范围内变化，因为铅 - 210 蜕变的原子数同它存在的数量成正比．这样一来，我们还是不能利用(9) 得到油画年龄的精确估值，甚至也不能得到粗略的估值．

表 15 - 9　矿石和精选矿石样品(所有蜕变率按每克铅白计算)

矿石种类和产地	每分钟 Ra^{226} 蜕变的原子数
精选矿石(俄克拉何马 - 堪萨斯)	4.5
破碎的原矿石(密苏里东南部)	2.4
精选矿石(密苏里东南部)	0.7
精选矿石(爱达荷)	2.2
精选矿石(爱达荷)	0.18
精选矿石(华盛顿)	140.0
精选矿石(不列颠哥伦比亚)	1.9
精选矿石(不列颠哥伦比亚)	0.4
精选矿石(玻利维亚)	1.6
精选矿石(澳大利亚)	1.1

我们需要换一种方法利用(9) 来区别 17 世纪的油画和现代的赝品．这是根据下述的简单事实：如果颜料的年头比起铅的半衰期 22 年老得多，那么颜料中铅 - 210 的放射量几乎等于颜料中镭的放射量．另一方面，如果油画是现代作品，譬如是 20 年上下的作品，那么铅 - 210 的放射量就比镭的放射量大得多．

我们把这一论点作如下的精确分析．所考察的油画或者是很新的，或者有 300 年之久．在(9) 中令 $t - t_0 = 300$(年)，我们有，

$$y(t) = \frac{r}{\lambda}(1 - e^{-300\lambda}) + y_0 e^{-300\lambda},$$

$$\lambda y(t) = r(1 - e^{-300\lambda}) + \lambda y_0 e^{-300\lambda}$$

于是，

$$\lambda y_0 = \lambda y(t) e^{300\lambda} - r(e^{300\lambda} - 1). \tag{10}$$

如果这帧油画是古代作品，λy_0 就不大，如果它的确是现代的伪造品，那么就会大得出奇．为了确定怎样才算大得出奇，我们注意到，如果在当初制造颜料时，每克铅白中所含铅－210 每分钟的蜕变数是 100 个原子，那么提炼出它的那种矿石大约含 0.014％ 的铀．这个铀的浓度是相当高的，因为地壳岩石中铀的平均含量约为 2.7×10^{-6}．另一方面，在西半球有些极罕见的矿石中含铀量达 2％～3％．为保险起见，如果每克铅白中所含铅－210 的蜕变率超过每分钟 30 000 个，那么这样的蜕变率就肯定是大得出奇．

要计算 λy_0，必须计算此刻铅－210 的蜕变率 λy，镭－226 的蜕变率 r 和 $e^{300\lambda}$．因为钋－210（Po^{210}）的蜕变率等于铅－210 若干年后的蜕变率，而钋－210 的蜕变率较容易测量，所以我们用钋－210 的蜕变率代替铅－210 的蜕变率．由(4)，$\lambda = \frac{\ln 2}{22}$，因此，

$$e^{300\lambda} = e^{300 \ln 2/22} = 2^{150/11}.$$

对于油画“埃牟斯的门徒”和其它几帧疑为伪制品的画，测量出钋－210 和镭－226 的蜕变率，列在表 15－10 中．

表 15－10　作者有疑问的油画

（所有蜕变率按每克铅白每分钟计算）

油画名称	Po^{210} 蜕变的原子数	Ra^{226} 蜕变的原子数
埃牟斯的门徒	8.5	0.8
濯　足	12.6	0.26
看乐谱的女人	10.3	0.3
演奏曼陀琳的女人	8.2	0.17
花边织工	1.5	1.4
笑　女	5.2	6.0

现在,如果我们对于油画“埃牟斯的门徒”中的铅白,由(10)算出之值,就得到

$$\begin{aligned}\lambda y_0 &= 8.5 \times 2^{150/11} - 0.8(2^{150/11} - 1) \\ &= 98\,050.\end{aligned}$$

这个数大得难以置信.因此,这幅画一定现代的伪制品.类似地分析无可争辩的证明了,油画“濯足”,“看乐谱的女人”,“演奏曼陀琳的女人”都不是弗美尔的作品.另一方面,“花边织工”和“笑女”都不可能是现代的伪制品,因为在这两幅画中,钋-210和镭-226都非常接近放射性平衡.

小结

1. 预备知识.这里需要的知识主要是微积分.要熟悉微分法和基本积分法,对简单的微分方程及其解法要有一些经验.

2. 一个好的模型需要满足两个条件:一是,用到的假设要尽量地少,用到的数学要尽量地简单;二是,它不仅对过去的观察现象作出精确的描述,而且要对未来作出预测.

3. 构造数学模型在整个建模过程中扮演了主角.典型的建模过程包括将实际问题理想化、近似化.这就是用一个更简单的,但却保持了原问题本质特征的问题去代替原问题.建模的要点就是选择一种合适的数学结构,使实际问题中的概念和关系与数学系统中的概念和关系有效地对应起来.

开始建模时,常用一个简单的模型来研究问题.其好处是,简单模型易于作数学处理.然后在此基础上,每次增加一个或几个新的假设,使模型逐步完备起来.在人口问题中,我们选取的第一个模型是

$$\frac{\mathrm{d}y}{\mathrm{d}t} = ky.$$

这个模型与过去的观测值有很好的近似,是一个相当好的模型.这个

模型的缺点是，它不能解决人口数量的长期预测．于是我们又增加了环境因素对人口增长的影响，而得到第二个模型．第二个模型满足了世界人口的数量不能无限增长的事实，比第一个模型更加合理．

所以建模过程是一个逐渐演化的过程，又是一个逐渐优化的过程．一般说来，任何模型都不可能包罗万象．所以有时有几个模型，每个模型说明问题的某些方面．

4．具体过程．一个实际问题常常涉及到一个量 y，它是时间 t 的函数．要求建立一个联系 y,y' 和时间 t 的方程，它在任何时刻都成立．对这个方程积分就得到一个只含有 y 和 t，而不含 y' 的新方程，这个新方程中含有任意常数，并且对任意特定的 t 仍然成立．这就是方程的通解．实际问题中给出了在特定时刻成立的信息，这个信息用来计算积分常数，或其它参数．最后，我们得到函数 $y(t)$．

5．注意事项．

1）转化．在实际问题中有许多表示“导数”的常用词，如“速率”、“速度”、“衰变”等．当实际问题中出现“改变”、“变化”、“增加”、“减少”时，就是可能出现导数的信号．所以，转化的第一步是把实际问题中的概念变为数学概念．

转化的第二步是实际问题中的规律，如物理定律，生物定律，经济规律等转化为数学方程，这种方程常常是微分方程．

2）微分方程．微分方程是一个在任何时刻都必须正确的瞬时表达式．这是数学问题的核心．下面的模式是常常用到的：

$$纯变化率 = 输入率 - 输出率.$$

3）单位．一旦确定了哪些项进入微分方程中，你必须确保每一项都采用相同的单位．这样才能保证最后的工作是可用的．

4）给定的条件．这是系统在某一个特定时刻的信息，它们独立于微分方程．在微分方程解出后，利用它们去确定有关的常数，其中包括比例常数，积分常数，以及微分方程中的其它参数．

5）框架．把你解题的总框架写下来，这能帮助你抓住整个解题

过程中的主要步骤.并且,别人也能看懂你的工作.主要步骤是:

转化.实际概念转化为数学概念;有关定律转化为数学方程.

建立微分方程.

选好单位.

写出给定的条件.

解数学问题.

回到实际问题中去.

习 题

1. 铀 − 238 经过一系列过程蜕变为铅 − 206,不再具有放射性.铀 − 238 的半衰期是 4.5×10^9 年.现在有一块岩石,经分析它含有铀 − 238,也含有铅 − 206.铀的含量是铅的 2 倍.问这块岩石有多老?

2. 镭的半衰期是 1622 年,问蜕变出 25% 需要多长时间?

3. 将室内一支读数为华氏 60 度的温度计放到室外.10 分钟后读数为 70 度.又过了 10 分钟读数为 76 度.先不用计算,推测一下室外的温度.然后用牛顿冷却定律计算出正确的答案.

第十六章　外微分形式

一种好的记号可以使头脑摆脱不必要的负担和约束，使思想集中于新的问题；这就事实上增加了人脑的能力.

A.H.Whitehead

数学公式有其自身的独立存在性与智慧，它们比我们聪明，甚至比它们的发现者也聪明，并且我们从它们中得到的比原来注入的要多.

H.Hertz

这部分的内容实质上属于流形上的微积分.我们只介绍一些最初步的知识，以求扩大眼界.因而我们的着眼点不是数学上的严格，而是希望对外微分形式有一个较为直观的了解.由此，对微积分的内容可获得一个比较深入的本质的认识，弄清楚在高维空间中微分和积分这对矛盾是如何体现的.

在一元微积分中，牛顿－莱布尼茨公式是最重要的公式，它建立了微分学和积分学之间的联系.在多元微积分中它的类似物是什么呢?我们就来回答这一问题.在引入外微分形式后，就会看得很清楚.

外微分形式是一种很好的数学符号和运算，用它可以统一处理多元微积分的有关内容，如场论中的三个公式及积分换元法.

16.1.1　场论的三个基本公式

在场论中有三个重要的基本公式，它们将线、面、体积分联系了起来.这三个公式是

1）格林公式

$$\oint_L P\mathrm{d}x + Q\mathrm{d}y = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y,$$

这里 D 是一个平面区域，L 是它的边界．P，Q 都是 x，y 的函数，在 D 内具有连续的一阶偏导数．

2）高斯公式

$$\oiint_S P\mathrm{d}y\mathrm{d}z + Q\mathrm{d}z\mathrm{d}x + R\mathrm{d}x\mathrm{d}y$$
$$= \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

这里 V 是由闭曲面 S 围成的空间区域．P，Q，R 是 x，y，z 的函数，在区域 V 内具有连续的一阶偏导数．

3）斯托克斯公式

$$\oint_L P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z$$
$$= \iint_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathrm{d}y\mathrm{d}z + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathrm{d}z\mathrm{d}x$$
$$+ \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y$$

这里闭曲线 L 是曲面 S 的边界，P，Q，R 具有连续的一阶偏微商．

这三个公式有联系吗？能不能用一个统一的公式将它们概括起来？可不可以将其推广到更高维的情况？

16.1.2 曲面的定向

我们在学习曲线积分和曲面积分的时候已经注意到，线积分、面积分的区域都是有方向的（指第二种曲线积分和曲面积分，以下都是如此）．一重积分和二重积分可以视为曲线积分和曲面积分的特殊情形，因而它们的积分区域也是有方向的．同样，对三重积分的区域也可以定向．

为了对区域的定向有个初步的了解，我们以曲面为例，稍作说明．

一般说来，一张平面或一个曲面都有两侧.例如，一张平放着的平面有上下两侧.再如，闭球面有内外两侧，等等.

怎样给一张平面定向呢？我们指定一侧为正，则另一侧为负.通常我们在平面上选定一侧为正侧，并用它的法线方向来表示这一侧.然后，在其上画出直角坐标系.伸出你的右手，如果从 x 轴向 y 轴握手，你的大拇指指向平面的法线方向，那么这个坐标系就称为右手系，否则，称为左手系.因而，我们可以用右手系代表平面的正侧(图 16－1).

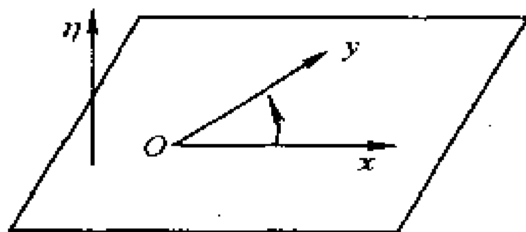


图 16－1



图 16－2

对三维欧氏空间，我们也用空间右手坐标系来规定它的正向.

对曲面来说，我们用曲面在一点的法方向来表示曲面的正向.

是不是所有的曲面都有两侧呢？不是的.确有这样的曲面，它只有一侧，例如莫比乌斯曲面就是单侧曲面(图 16－2).莫比乌斯带是这样作成的：取一条纸带，把它扭半转，然后将两端粘起来.

16.1.3　外乘积

通过平面和空间的例子，我们引进外乘积的概念.

先看平面情况.设 $a_1=(a_{11},a_{12})$，$a_2=(a_{21},a_{22})$ 是两个平面向量.我们在这两个向量之间定义一种运算：

$$a_1 \wedge a_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

它有明显的几何意义：表示由向量 a_1，a_2 所张成的平行四边形的有

向面积(图 16 − 3)."有向"是指,这个面积具有正负号.如果改变向量 a_1, a_2 的次序,面积就改变符号.借助行列式的性质,容易验证运算"∧"具有下面的性质:

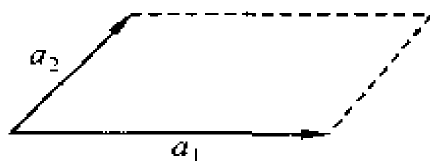


图 16 − 3

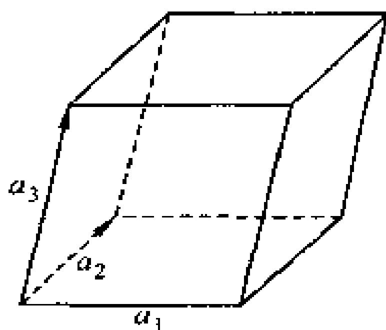


图 16 − 4

1) 反交换性　$a \wedge b = -b \wedge a$;

2) 线性　$a \wedge (b + c) = a \wedge b + a \wedge c$,

$(\lambda a) \wedge b = a \wedge (\lambda b) = \lambda(a \wedge b)$　(λ 是实数).

再看空间的情形.设 $a_i = (a_{i1}, a_{i2}, a_{i3})(i = 1,2,3)$ 是空间中的三个向量.我们定义:

$$a_1 \wedge a_2 \wedge a_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

它也有明显的几何意义,就是由向量 a_1, a_2, a_3 所张成的平行六面体的有向体积(图 16 − 4).借助行列式的性质,容易验证这里定义的运算"∧"也具有反交换性和线性的性质.

有了这些直观背景后,我们引入外乘积的概念.

如果在空间中(我们只考虑 $n = 2,3$ 的情况)定义了一种运算"∧",使得任意 k 个向量 $a_{1,2}, \cdots, a_k$ 具有下面两条性质:

1) 反交换性　$\cdots \wedge a_e \wedge a_f \wedge \cdots = -(\cdots \wedge a_f \wedge a_e \wedge \cdots)$;

2) 线性　　$\cdots \wedge (\lambda a_e + \mu a_f) \wedge \cdots = \lambda(\cdots \wedge a_e \wedge \cdots) + \mu(\cdots \wedge a_f \wedge \cdots)$,

我们就称这一运算为外乘积,这里 λ,μ 是实常数.

例　设 e_1,e_2 是平面上两个线性无关的向量,因而可以看成平面上的一组基.于是平面上任意两个向量 a_1,a_2 可以用 e_1,e_2 线性表出:

$$a_1 = a_{11}e_1 + a_{12}e_2,\quad a_2 = a_{21}e_1 + a_{22}e_2.$$

利用外乘积的性质,我们首先注意到

$$e_1 \wedge e_2 = - e_2 \wedge e_1,$$

由此自然有,

$$e_1 \wedge e_1 = e_2 \wedge e_2 = 0.$$

于是我们有下面的计算

$$\begin{aligned} a_1 \wedge a_2 &= (a_{11}e_1 + a_{12}e_2) \wedge (a_{21}e_1 + a_{22}e_2) \\ &= a_{11}e_1 \wedge (a_{21}e_1 + a_{22}e_2) + a_{12}e_2 \wedge (a_{21}e_1 + a_{22}e_2) \\ &= a_{11}a_{22}e_1 \wedge e_2 + a_{12}a_{21}e_2 \wedge e_1 \\ &= a_{11}a_{22}e_1 \wedge e_2 - a_{12}a_{21}e_1 \wedge e_2 \\ &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} e_1 \wedge e_2. \end{aligned}$$

注意到 $a_1 \wedge a_2$ 和 $e_1 \wedge e_2$ 的几何意义,我们可将上面的公式用几何语言表达出来:由向量 a_1,a_2 所张成的平行四边形的有向面积与由向量 e_1,e_2 所张成的平行四边形的有向面积的比是

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

这个行列式是面积的放大系数.当行列式大于零时,向量 a_1,a_2 与 e_1,e_2 有相同的定向;当行列式小于零时,向量 a_1,a_2 与 e_1,e_2 有相反的定向(图 16 − 5).

这个例子启发我们去考虑二重积分的换元公式.

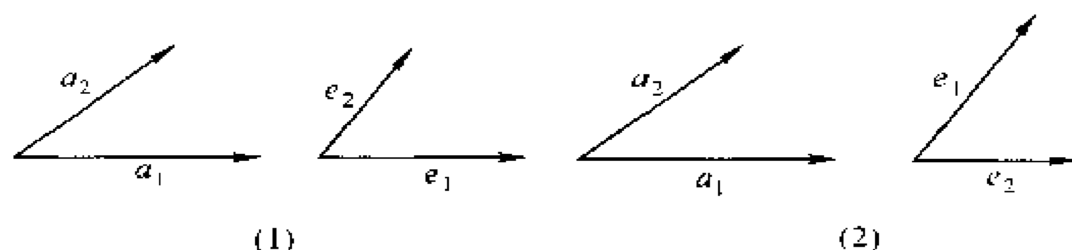


图 16 · 5

例　设 e_1,e_2,e_3 是空间中三个线性无关的向量，a_1,a_2,a_3 是空间中另外三个任意向量. 于是向量 a_1,a_2,a_3 可以用 e_1,e_2,e_3 线性表出：

$$a_i = a_{i1}e_1 + a_{i2}e_2 + a_{i3}e_3 \quad (i = 1,2,3).$$

利用外乘积的性质可以算出(具体计算留给读者)：

$$a_1 \wedge a_2 \wedge a_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} e_1 \wedge e_2 \wedge e_3.$$

右端的行列式是体积放大系数.

现在我们来研究它们在重积分中的应用. 考虑二重积分 $\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$. 作变量替换

$$x = x(u,v),$$
$$y = y(u,v).$$

我们假定变换函数有连续的一阶偏导数. 求出 x 和 y 的全微分

$$\mathrm{d}x = \frac{\partial x}{\partial u}\mathrm{d}u + \frac{\partial x}{\partial v}\mathrm{d}v, \mathrm{d}y = \frac{\partial y}{\partial u}\mathrm{d}u + \frac{\partial y}{\partial v}\mathrm{d}v.$$

注意到，$\mathrm{d}x\mathrm{d}y$ 是由 $\mathrm{d}x$ 和 $\mathrm{d}y$ 所张成的小平行四边形(实际上是矩形)的面积，所以这里出现的乘积是外乘积. 记着，

$$\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x; \quad \mathrm{d}x \wedge \mathrm{d}x = \mathrm{d}y \wedge \mathrm{d}y = 0,$$

作外乘积，我们得到，

$$\mathrm{d}x \wedge \mathrm{d}y = \left(\frac{\partial x}{\partial u}\mathrm{d}u + \frac{\partial x}{\partial v}\mathrm{d}v\right) \wedge \left(\frac{\partial y}{\partial u}\mathrm{d}u + \frac{\partial y}{\partial v}\mathrm{d}v\right)$$

$$= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \mathrm{d}u \wedge \mathrm{d}v = \frac{\partial(x,y)}{\partial(u,v)}\mathrm{d}u \wedge \mathrm{d}v,$$

这个行列式就是我们所熟知的雅可比行列式.当行列式大于0时,xy平面与uv平面有相同的定向;当行列式小于0时,xy平面与uv平面有相反的定向.

我们记得,积分变换公式是这样的

$$\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \iint_D f(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\mathrm{d}u\mathrm{d}v,$$

其中行列式取了绝对值,这是因为两个坐标系都取的是正定向,为了保证面积元素是正的,所以取绝对值.但是在已经定向的曲面上积分时,面积本来可正可负,因此就没有必要对雅可比行列式取绝对值了.此时

$$\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \iint_D f(x(u,v),y(u,v))\frac{\partial(x,y)}{\partial(u,v)}\mathrm{d}u\mathrm{d}v.$$

例　求极坐标下的雅可比行列式.

解　由 $x = r\cos\theta, y = r\sin\theta$,得

$$\mathrm{d}x = \cos\theta\mathrm{d}r + r\,\mathrm{d}\cos\theta = \cos\theta\mathrm{d}r - r\sin\theta\mathrm{d}\theta$$
$$\mathrm{d}y = \sin\theta\mathrm{d}r + r\,\mathrm{d}\sin\theta = \sin\theta\mathrm{d}r + r\cos\theta\mathrm{d}\theta.$$

注意到,$\mathrm{d}r \wedge \mathrm{d}\theta = -\mathrm{d}\theta \wedge \mathrm{d}r, \mathrm{d}r \wedge \mathrm{d}r = \mathrm{d}\theta \wedge \mathrm{d}\theta = 0$,我们有

$$\mathrm{d}x \wedge \mathrm{d}y = (\cos\theta\mathrm{d}r - r\sin\theta\mathrm{d}\theta) \wedge (\sin\theta\mathrm{d}r + r\cos\theta\mathrm{d}\theta)$$
$$= r\cos^2\theta\mathrm{d}r \wedge \mathrm{d}\theta - r\sin^2\theta\mathrm{d}\theta \wedge \mathrm{d}r$$

$$= r(\cos^2\theta + \sin^2\theta)\mathrm{d}r \wedge \mathrm{d}\theta = r\mathrm{d}r \wedge \mathrm{d}\theta$$

这就算出了

$$\frac{\partial(x,y)}{\partial(r,\theta)} = r.$$

这个结果是大家熟知的.

三重积分的情形也一样. 考虑三重积分 $\iiint_V f(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$. 作变量替换

$$x = x(u,v,w),$$
$$y = y(u,v,w),$$
$$z = z(u,v,w).$$

我们假定变换函数有连续的一阶偏导数. 求出 x,y,z 的全微分

$$\mathrm{d}x = x_u\mathrm{d}u + x_v\mathrm{d}v + x_w\mathrm{d}w,$$
$$\mathrm{d}y = y_u\mathrm{d}u + y_v\mathrm{d}v + y_w\mathrm{d}w,$$
$$\mathrm{d}z = z_u\mathrm{d}u + z_v\mathrm{d}v + z_w\mathrm{d}w,$$

因而

$$\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z = \begin{vmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{vmatrix} \mathrm{d}u \wedge \mathrm{d}v \wedge \mathrm{d}w$$
$$= \frac{\partial(x,y,z)}{\partial(u,v,w)}\mathrm{d}u \wedge \mathrm{d}v \wedge \mathrm{d}w.$$

这就是三重积分变量变换中体积元素的表达式. 这里的雅可比行列式是有正负号的. 其符号表明了两个空间的定向.

n 重积分也有类似的公式.

16.1.4 微分形式和它的外微分

任何积分运算实质上都是微分运算, 因而对微分式作特别的讨论是必要的. 以下的讨论限定在三维空间.

设 A,B,C,H,P,Q,R 为 x,y 的函数, 或 x,y,z 的函数. 我们

称函数

$$\omega_0 = f(x, y, z)$$

为零次微分形式，称

$$\omega_1 = A\mathrm{d}x + B\mathrm{d}y + C\mathrm{d}z$$

为一次微分形式，称

$$\omega_2 = P\mathrm{d}y \wedge \mathrm{d}z + Q\mathrm{d}z \wedge \mathrm{d}x + R\mathrm{d}x \wedge \mathrm{d}y$$

为二次微分形式，称

$$\omega_3 = H\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$$

为三次微分形式.

对二维情况，

$$\omega_0 = f(x, y)$$

$$\omega_1 = A\mathrm{d}x + B\mathrm{d}y$$

$$\omega_2 = P\mathrm{d}x \wedge \mathrm{d}y$$

例　下面是几个微分形式的例子：

$x\mathrm{d}x + \sin x\mathrm{d}y + y^2\mathrm{d}z$；$z\mathrm{d}x \wedge \mathrm{d}y + x\mathrm{d}y \wedge \mathrm{d}z + y\mathrm{d}z \wedge \mathrm{d}x$；$(u + v + w)\mathrm{d}u \wedge \mathrm{d}v \wedge \mathrm{d}w$.

我们在微分形式上定义外微分运算 d 如下：

1）零次微分形式，在 2 维情况下，$f = f(x, y)$，定义

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y$$

在 3 维情况下，$f = f(x, y, z)$，定义

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y + \frac{\partial f}{\partial z}\mathrm{d}z.$$

这就是说，零次微分形式的外微分运算就是普通的全微分算子，其结果是一个一次微分形式.

例　$u = x^2 + y^2$，$\mathrm{d}u = \mathrm{d}x^2 + \mathrm{d}y^2 = 2x\mathrm{d}x + 2y\mathrm{d}y$.

$u = xyz$，$\mathrm{d}u = \mathrm{d}(xyz) = yz\mathrm{d}x + xz\mathrm{d}y + xy\mathrm{d}z$.

2）对一次微分形式. 在 2 维情况，设 $\omega = P\mathrm{d}x + Q\mathrm{d}y$，定义

$$
\begin{aligned}
\mathrm{d}\omega &= (\mathrm{d}P) \wedge \mathrm{d}x + (\mathrm{d}Q) \wedge \mathrm{d}y \\
&= \left(\frac{\partial P}{\partial x}\mathrm{d}x + \frac{\partial P}{\partial y}\mathrm{d}y\right) \wedge \mathrm{d}x + \left(\frac{\partial Q}{\partial x}\mathrm{d}x + \frac{\partial Q}{\partial y}\mathrm{d}y\right) \wedge \mathrm{d}y \\
&= \frac{\partial P}{\partial y}\mathrm{d}y \wedge \mathrm{d}x + \frac{\partial Q}{\partial x}\mathrm{d}x \wedge \mathrm{d}y \\
&= \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y.
\end{aligned}
$$

注意,这正是格林公式中的右边积分的被积项.

在 3 维情况下,

$$\omega_1 = A\mathrm{d}x + B\mathrm{d}y + C\mathrm{d}z.$$

外微分的相应定义为

$$
\begin{aligned}
\mathrm{d}\omega_1 &= (\mathrm{d}A) \wedge \mathrm{d}x + (\mathrm{d}B) \wedge \mathrm{d}y + (\mathrm{d}C) \wedge \mathrm{d}z \\
&= (A_x\mathrm{d}x + A_y\mathrm{d}y + A_z\mathrm{d}z) \wedge \mathrm{d}x \\
&\quad + (B_x\mathrm{d}x + B_y\mathrm{d}y + B_z\mathrm{d}z) \wedge \mathrm{d}y \\
&\quad + (C_x\mathrm{d}x) + C_y\mathrm{d}y + C_z\mathrm{d}z) \wedge \mathrm{d}z \\
&= (B_x - A_y)\mathrm{d}x \wedge \mathrm{d}y \\
&\quad + (C_y - B_z)\mathrm{d}y \wedge \mathrm{d}z \\
&\quad + (A_z - C_x)\mathrm{d}z \wedge \mathrm{d}x \\
&= \begin{vmatrix} \mathrm{d}y \wedge \mathrm{d}z & \mathrm{d}z \wedge \mathrm{d}x & \mathrm{d}x \wedge \mathrm{d}y \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A & B & C \end{vmatrix}
\end{aligned}
$$

这是一个二次微分形式,正是斯托克斯公式下右边积分的被积项.

3) 二次微分形式

$$\omega_2 = A\mathrm{d}y \wedge \mathrm{d}z + B\mathrm{d}z \wedge \mathrm{d}x + C\mathrm{d}x \wedge \mathrm{d}y$$

的外微分定义为

$$
\begin{aligned}
\mathrm{d}\omega_2 &= (\mathrm{d}A)\mathrm{d}y \wedge \mathrm{d}z + (\mathrm{d}B)\mathrm{d}z \wedge \mathrm{d}x + (\mathrm{d}C)\mathrm{d}x \wedge \mathrm{d}y \\
&= (A_x\mathrm{d}x + A_y\mathrm{d}y + A_z\mathrm{d}z)\mathrm{d}y \wedge \mathrm{d}z
\end{aligned}
$$

$$
\begin{aligned}
&+ (B_x \mathrm{d}x + B_y \mathrm{d}y + B_z \mathrm{d}z)\mathrm{d}z \wedge \mathrm{d}x \\
&+ (C_x \mathrm{d}x + C_y \mathrm{d}y + C_z \mathrm{d}z)\mathrm{d}x \wedge \mathrm{d}y \\
&= (A_x + B_y + C_z)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z,
\end{aligned}
$$

这是一个三次微分形式. 三次微分形式

$$\omega_3 = H\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$$

的外微分定义为,

$$
\begin{aligned}
\mathrm{d}\omega_3 &= (\mathrm{d}H) \wedge \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \\
&= (H_x \mathrm{d}x + H_y \mathrm{d}y + H_z \mathrm{d}z) \wedge \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z = 0,
\end{aligned}
$$

即

$$\mathrm{d}\omega_3 = 0.$$

这是因为每一项中至少有两个微分是相同的. 因而在三维空间中任何三次微分形式的外微分都是 0. 在这些规定下,外微分算子 d 与普通微分算子是一样的,即对每一项进行运算,在每一项中分别对每一个因子进行运算,其余因子不动,然后将所得的各项相加. 不同的只是外微分算子 d 在运算之后进行外乘,而普通微分算子是运算之后进行通常的乘积.

16.1.5　在场论中的应用

场论中的三个公式可以用微分形式和它们的外微分作统一处理. 由此我们就可以说清楚在高维空间中微分和积分如何成为一对矛盾了. 先看格林公式

$$\oint_L P\mathrm{d}x + Q\mathrm{d}y = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y.$$

如果我们取一次微分形式

$$\omega = P\mathrm{d}x + Q\mathrm{d}y,$$

则它的外微分是

$$\mathrm{d}\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y.$$

于是格林公式可以写成

$$\oint_{\partial D} \omega = \iint_D \mathrm{d}\omega .$$

线积分的曲线 L 是区域 D 的定向边界，我们用 ∂D 来表示(图 16－6)．

再看高斯公式

$$\oiint_S P\mathrm{d}y\mathrm{d}z + Q\mathrm{d}z\mathrm{d}x + R\mathrm{d}x\mathrm{d}y$$

$$= \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z .$$

我们取二次微分形式

$$\omega = P\mathrm{d}y \wedge \mathrm{d}z + Q\mathrm{d}z \wedge \mathrm{d}x + R\mathrm{d}x \wedge \mathrm{d}y ,$$

它的外微分是

$$\mathrm{d}\omega = \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z ,$$

于是高斯公式可以写成

$$\oiint_{\partial V} \omega = \iiint_V \mathrm{d}\omega .$$

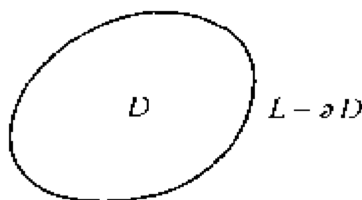


图 16－6

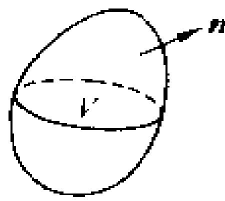


图 16－7

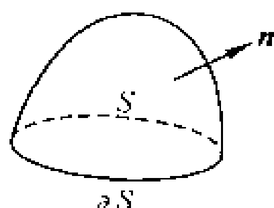


图 16－8

∂V 是包围 V 的闭曲面，其定向如图 16－7 所示．

最后看斯托克斯公式

$$\oint_L P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z = \iint_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathrm{d}y\mathrm{d}z$$

$$+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\mathrm{d}z\mathrm{d}x+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y$$

取一次微分形式

$$\omega = P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z,$$

它的外微分是

$$\begin{aligned}\mathrm{d}\omega = &\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\mathrm{d}y \wedge \mathrm{d}z\\ &+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\mathrm{d}z \wedge \mathrm{d}x\\ &+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y,\end{aligned}$$

于是斯托克斯公式可以写为

$$\oint_{\partial S}\omega = \iint_S \mathrm{d}\omega,$$

其中 ∂S 是曲面 S 的定向边界(图 16－8).

这样一来,场论的三个基本公式统一成了一个共同的形式

$$\int_{\partial D}\omega = \int_D \mathrm{d}\omega.$$

我们称这个公式为广义的斯托克斯公式,其中 ∂D 是 D 的边界.$\int$ 表示区域的维数是多少,积分的重数就是多少.

上面的公式可以推广到高维空间中去,可以推广到更一般的流形上去.这个一般的公式揭示了在高维空间中微分和积分是如何成为一对矛盾的.这对矛盾的一方是外微分形式 $\mathrm{d}\omega$,另一方是线、面、体积分,高次的外微分形式 $\mathrm{d}\omega$ 在某区域上的积分等于低一次的外微分形式 ω 在该区域的低一维的空间的边界上的积分.外微分运算与积分起了相互抵消的作用.在一维空间中就是微积分基本定理

$$\int_a^b \mathrm{d}f = f(b) - f(a)$$

D 就是这里的区间$[a,b]$,∂D 就是端点 a 和 b.

习　题

1 计算下列外乘积：

1) $(dx + 3dy) \wedge (2dx + 4dy)$;

2) $(x^2 dx + z dy) \wedge (2y dx - \cos x dy + z dz)$;

3) $(dx \wedge dy + 3dx \wedge dz) \wedge (dx + 2dy + dz)$.

2 计算下列外微分：

1) $d(\sin y dx + \cos x dy)$;

2) $d(z^2 dx \wedge dy - xy dx \wedge dz)$.

第十七章　数学的真理性

数学的终极基础和终极意义尚未解决,我们不知道沿着什么方向可以找到最终答案,或者甚至于是否有希望找到一个最终的、客观的答案.

H.魏尔

…至今已有25个世纪之久的这段时期,数学家们一直在改正他们的错误,而且看到了这门学科欣欣向荣,而不是枯竭衰败.这使他们有权力对未来充满希望.

布尔巴基

我们必须知道,我们将会知道.

希尔伯特

17.1.1　数学的证明与科学的证明

数学的证明与科学的证明之间存在着深刻的差别.这种差别是理解自毕达哥拉斯以来每个数学家工作的关键之点.

经典数学的证明方法是,从一系列公理、定义出发,通过逻辑论证,一步一步地得到某个结论.如果公理是正确的,逻辑又没有缺陷,那么得到的结论将是不可否定的.这个结论就是一个定理.

数学证明依靠这一逻辑过程,而且一个定理一经证明就永远是对的.为了正确判断这种证明的价值,应该将它们与科学证明作一比较.在物理学中,一个假设被提出来,用以解释某一类物理现象.如果对物理现象的观察与这个假设相符,就成为这个假设成立的证据.进而,这个假设不仅能描述已知的现象,而且能预见新的结果.如果它再次成功,那么就有更多的证据支持这个假设.最终,证据的数量可能达到压倒的程度,这个假设就作为一个理论而被接受.

科学的证明依赖于观察、实验和理解力，而这两者都是容易出错的，从而它只能提供近似真理的概念．即使人们最为普遍地接受了的科学证明中也总存在着可疑的成分．而在另外一些场合，这种理论最终会被证明是错的，这就导致科学上的革命，用一种新的理论去替代原以为正确的旧理论．这种新理论可能是原有理论的深化，也可能与原有理论完全相反．例如对物质基本粒子的探索，使得每一代物理学家都推翻或重新修改他们前辈的理论．

数学证明却与此不同．数学证明具有绝对的意义，是无可怀疑的．毕达哥拉斯公元前 500 年证明的定理，今天依然正确．数学不依赖于容易出错的实验证据，而是立足于逻辑．

数学证明与科学证明有着质的不同．但是，数学真理就是绝对真理吗？公理化方法的地位是稳如泰山，而无懈可击吗？

17.1.2 数学的公理化

虽然古希腊人对数学作出过许多贡献，但他们对此学科作出的最大贡献或许是他们用公理化方法对数学所作的整理．此后的两千年中，公理化方法没有得到进一步的发展．直到 19 世纪，数学发展中的三个重大事件才导致数学家们对公理化方法作深一步的研究．这三个重大事件是，非欧几何的发现，非交换代数的发现，以及以分析的算术化为顶峰的实数理论的建立．

由于 19 世纪许多数学家的工作，公理化的方法逐渐变得清晰而精细．希尔伯特的《几何基础》对欧几里得几何公设作了通俗而又严格的发展，于 1899 年发表，为重建公理化方法作出了重大贡献．希尔伯特把公理化思想明确而严格地确立了下来．他对公理化提出了一些逻辑上的要求：

1）完备性．完备性指，所有的定理都可以从这组公理中推导出来．

完备性的要求出现得很自然，每门科学都有这一要求，因为人们总希望科学能回答一切问题，数学也不例外．负数的引进，无理数的

引进，以及复数的引进都是为了满足完备性的要求.若一组公理是完备的，那么所有的定理就都能从这组公理中推导出来.要是从中去掉一个公理，一些定理将得不到证明.

2）独立性.称一组公理中的一个公理是独立的，如果它不是其它公理的逻辑推论；整个公理组是独立的，如果它的每一个公理都是独立的.

数学史上关于公理的独立性的最著名的研究是关于欧几里得的平行公设的研究.前面我们曾做过详细介绍.我们知道，最后证明欧几里得的平行公设的独立性的是罗巴切夫斯基非欧几何的发现及其相容性的证明.

独立性的证明不是绝对必要的.一组公理显然不会因为缺少独立性就成为无用的.但数学家们偏爱独立性，因为他们要把他们的理论建立在最少的假定之上.一组不独立的公理集自然不满足这一要求.

有些著名的公设集在最初发表时，人们并不知道它包含了不独立的公设.例如，希尔伯特起初为欧几里得几何安排的公设集就是如此.后来发现，这组公设集中有两条公设就蕴含在其它公设中.这两条不独立的公设的发现并没有使希尔伯特的公设失去效力，只不过把这两条公设变成定理罢了.

3）相容性.其含义是，从这些公理出发不能推出相互矛盾的定理来.这是一组公理集的最重要、最基本的性质.没有这条性质，这组公理集就毫无价值.

17.1.3　天衣有缝

微积分的发现把无限引入了数学，同时也引出了第二次数学危机.到19世纪末，分析的严格化问题得到了解决.柯西建立了严格的极限理论，魏尔斯特拉斯引进了 $\epsilon-\delta$ 语言，戴德金、康托尔等又将实数理论严密化.分析有了可靠的基础和完整的体系.第二次数学危机终于过去了.这样，1900年在巴黎举行的第二次国际数学家大会上，

庞加莱高兴地指出：

"我们最终达到了绝对的严密吗？在数学发展前进的每一阶段，我们的前人都坚信他们达到了这一点.如果他们被蒙蔽了，我们是不是也像他们一样被蒙蔽了？…… 如果我们不厌其烦地严格的话，就会发现只有三段论或归结为纯数的直觉是不可能欺骗我们的.今天我们可以宣称，完全的严格性已经达到了！"

那时，绝大多数数学家具有和庞加莱相同的看法，他们为数学所达到的严密性而欢欣鼓舞.但实际上，暴风雨正在酝酿.屋外云涛翻滚，山雨欲来，数学史上的一场新的危机正在降临.

当时，还是有一些数学家已经清醒地认识到，数学基础中的漏洞并没有完全堵住.在这次会议上希尔伯特提出了他认为是数学发展中最重要的 23 个问题.希尔伯特的第一个问题就是康托尔的连续统基数问题，这引出了 1963 年美国数学家科恩的重要工作.在第二问题中他提出了相容性这个至关重要的问题.这两个问题都涉及到数学的基础是否稳固.可惜，许多数学家都没有意识到这个问题.即使希尔伯特也没有预见到，这个问题将会在怎样的广度和深度上席卷数学基础.

17.1.4 希尔伯特和他的 23 个问题

希尔伯特是 20 世纪最伟大的数学家之一，他于 1862 年 1 月 23 日生于德国的哥尼斯堡.他一直在家乡上学，1884 年获哥尼斯堡大学博士学位.1895 年担任著名的哥廷根大学的教授，1930 年退休，1943 年去世.他在数学的各个领域内都作出了具有深远影响的工作.

第二次国际数学家大会于 1900 年 8 月 6 日由著名数学家庞加莱宣布开幕.希尔伯特的演说本来是安排在开幕式上做的，由于他的迟疑而改在 8 月 8 日上午.这天上午，年仅 38 岁的希尔伯特登上了讲坛，发表了那篇影响深远的演说.这一演说成为数学史上的一个重要的里程碑，为 20 世纪的数学发展掀开了光辉的第一页.对 20 世纪数

学的发展起了巨大的推动作用.现在是20世纪的末期,21世纪即将来临.在世纪交替之际,回顾这篇演说具有特殊的意义.

希尔伯特朴素无华,他沉着的气度和卓越的才智吸引着每一个听众.他说:

"我们当中有谁不想揭开未来的帷幕,看一看在今后的世纪里我们这门学科发展的前景和奥秘呢?我们下一代的主要数学思潮将追求什么样的特殊目标呢?在广阔而丰富的数学思想领域,新世纪将会带来什么样的新方法和新成果?

历史教导我们,科学的发展具有连续性.我们知道,每一个时代都有它自己的问题,这些问题后来或者得以解决,或者因为无所裨益而被抛到一边,并为新的问题所替代.如果我们想对最近的将来数学知识的可能发展有一个概念,那就必须回顾一下当今科学提出的、期望在将来能够解决的问题.现在,当此世纪交替之际,我认为正适于对问题进行一番这样的检阅.因为,一个伟大时代的结束,不仅促使我们追溯过去,而且把我们的思想引向那未知的将来."

接着他指出,问题在数学发展中的重要性.他说:

"某类问题对于一般数学进展的深远意义以及它们在研究者个人的工作中所起的重要作用是不可否认的.只要一门科学分支中充满大量问题,它就充满了生命力;缺乏问题预示着独立发展的衰亡或终止.正如人类的每种事业都为了达到某种最终目的一样,数学研究也需要自己的问题.正是通过解决这些问题,研究者锻炼了自己的意志力,发现了新方法和新观点,达到更为广阔和自由的境界."

那么,如何判断一个问题有无价值呢?判断的标准是什么呢?他说:

"要想预先正确判断一个问题的价值是困难的,并且常常是不可能的,因为最终的判断取决于科学从该问题得到的收益.尽管如此,我们仍然要问,是否存在一般的准则,可用它鉴别出好的数学问题.一位法国老数学家曾经说过:'如果你不能向大街上遇到的第一个人

解释清楚你的数学理论,那么,你的理论就不能认为是完善的'.我以为,他对数学理论所坚持的清晰性和易懂性,应当作为一个堪称完善的数学问题的要求.因为,一个清晰易懂的问题会引起人们的兴趣,而复杂的问题使人们望而生畏.

其次,为了具有吸引力,一个数学问题应该是困难的,但又不应是完全不可解决的,而使我们劳而无功.在通向那隐藏的真理的曲折道路上,它应该是指引我们前进的一盏明灯,最终并以成功的喜悦作为我们的报偿."

简而言之,希尔伯特认为一个问题有价值,如果它满足两个条件:1)清晰易懂性;2)难而又可解决.我们前面介绍的许多著名问题都具有这一特色.例如,古代几何三大难题,费马大定理,哥德巴赫问题等都是.

特殊难题具有什么样的价值呢?希尔伯特说:

"以往的数学家习惯于以巨大的热情去致力于解决那些特殊的难题.他们懂得困难问题的价值.我只提醒大家注意伯努利提出的"最速降线问题".在公布这一问题时,伯努利说:'经验告诉我们,正是摆在面前的那些困难同时而又有用的问题,引导着有才能的人们为丰富人类的知识而奋斗'.…… 这个问题好比一块试金石,分析学家可以检验自己方法的价值,衡量他们的能力.变分法的起源应归功于伯努利的这一问题和类似的一些问题.

如所周知,费马曾断言丢番图方程

$$x^n + y^n = z^n (x, y, z \text{ 为整数})$$

除去某些自明的情况外是不可解的.证明这种不可解性的尝试提供了一个很好的例子,说明这样一个非常特殊、似乎不十分重要的问题会对科学产生怎样令人鼓舞的影响.受费马问题的启发,库默尔引进了理想数,并发现了把一个循环域的数分解为理想素因子的唯一分解定理,这一定理今天已由戴德金和克罗内克推广到任意代数域,并在近代数论中占有中心的地位,其意义远远超出数论的范围而深入

到代数和函数论的领域中去了.

说到另一个很不相同的研究领域,请大家注意三体问题.由庞加莱引进到天体力学中的那些卓有成效的方法和影响深远的原则,今天也被实用天文学家所确认和应用,而它们正是起因于庞加莱对三体问题的研究:他重新研究了这个困难问题并使它更接近解决.

……;

在回顾了问题在数学中的一般重要性之后,我们现在要转向这样一个问题:数学这门科学究竟以什么作为其问题的源泉呢?在每个数学分支中,那些最初、最老的问题肯定是起源于经验,是由外部的现象世界提出的.整数的运算法则就是以这种方式在人类文明的早期被发现的,与今天的儿童通过经验的方法来学习运用这些法则一样.对于最初的几何问题,诸如自古相传的立方倍积问题,化圆为方问题等等,情形也是如此.同样的还有数值方程的解、曲线论、微积分、傅里叶级数和位势理论中那些最初的问题,更不用说更大量的属于力学、天文学和物理学方面的问题了.

但是,随着一门数学分支的进一步发展,人类的智力由于受到成功的鼓舞,开始意识到自己的独立性.它自身独立地发展着,通常并不受来自外部的明显的影响,而是借助于逻辑组合,一般化和特殊化,巧妙地对概念进行分析和综合,提出新的富有成果的问题,因而它自己就以一个真正提问者的身份出现.这样就产生出素数问题以及伽罗瓦的方程式论、代数不变量理论、阿贝尔函数和自守函数等方面的一系列问题.确实,近代数论和函数论中几乎所有较深入的问题都是以这样的方式提出的.

其间,当纯思维的创造力进行工作时,外部世界又重新起作用,通过实际现象向我们提出新的问题,开辟新的数学分支.而当我们试图征服这些新的、属于纯思维王国的知识领域时,常常会发现过去未曾解决的问题的答案,这同时就极有效地推动着老的理论.据我看来,数学家们在他们这门科学各分支的问题提法、方法和概念中所经

常感觉到的那种令人惊讶的相似性和仿佛事先有所安排的协调性，其根源就在于思维和经验之间这种反复出现的相互作用.

下面，我想对在数学问题中常常遇到的困难和克服这些困难的办法作一些分析.

在解决一个数学问题时，如果我们没有获得成功，原因常常在于我们没有认识到更一般的观点，即眼下要解决的问题不过是一连串有关问题中的一个环节.采取这样的观点以后，不仅我们研究的问题会容易地得到解决，同时还会获得一种能应用于有关问题的普遍方法.柯西在定积分理论中引进复积分路径，库默尔在数论中引进“理想”的概念，就是这样的例子.这种寻求一般方法的途径肯定是最行得通的，也是最可靠的；手中没有明确的问题而去寻求一般方法的人，他们的工作多半是劳而无功的.

在讨论数学问题时，我们相信特殊化比一般化起着更为重要的作用.可能在大多数场合，我们寻求一个问题的答案而未能成功的原因是，有一些比手头的问题更简单、更容易的问题还没有完全解决，或完全没有解决.这时，一切都有赖于找出这些比较容易的问题，并使用尽可能完善的方法和能够推广的概念来解决它们.这种方法是克服数学困难的最重要的杠杆之一，我认为人们是经常使用它的，虽然也许并不自觉.

有时会碰到这种情况：我们是在不充分的前提下或不正确的意义上寻求问题的解答，因此不能获得成功.于是就产生了这样的任务：证明在所给的前提和所考虑的意义下原来的问题是不可解的.这样一种不可能性的证明古人就已实现.例如，他们证明了，等腰直角三角形的斜边与直角边的比是无理数.在后来的数学中，不可解的问题起着重要的作用.这样，我们领悟到：一些古老而困难的问题，诸如平行公理的证明，化圆为方，或用根式解五次方程等，已经得到令人满意和严格的解决，尽管是在与原先的企图不同的另一种意义上.

也许正是这一值得注意的事实加上其它哲学上的因素，给人们

以这样的信念(这信念为所有数学家所共有,但迄今没有一个人能给以证明):每一个数学问题都应该得到明确的解答.或者成功地解答所给问题,或者证明该问题不可解,即指出解答该问题的一切努力都将归于失败.……

这种相信每一个数学问题都可以解决的信念,对于数学家是一种巨大的鼓舞.我们之中常常听到这样的呼声:这里有一个数学问题,找出它的答案!你可以通过思维找出它的答案,因为数学中没有不可知.

数学问题的宝库是无穷无尽的,一个问题一旦解决,无数新的问题就代之而起.下面请允许我尝试着提出一些特定的问题,它们来源于数学的不同分支.通过这些问题的讨论,我们可以期待科学的进步."

下面是希尔伯特提出的23个问题:

1 证明'连续统假设',即证明任一实数集或者能与自然数集建立一一对应,或者能与全体实数集建立一一对应.

2 研究算术公理的相容性.

3 两个等底等高的四面体的体积相等.

4 直线作为两点间最短距离的问题.

5 李(S.lie)的连续群概念,不要定义群的函数的可微性假设.

6 物理学的公理化.

7 某些数的无理性和超越性.

8 素数问题.

9 在任意数域中证明最一般的互反定律.

10 丢番图方程的可解性.

11 系数为任意代数数的二次型.

12 阿贝尔域上的克罗内克定理在任意代数有理域上的推广.

13 不可能用仅有两个变数的函数解一般的七次方程.

14 证明某类完全函数系的有限性.

15 叔伯特(Schubert)计数演算的严格基础.

16 代数曲线与代数曲面的拓扑问题.

17 正定形式的平方和表示.

18 用全等多面体构造空间.

19 正则变分问题的解一定是解析的吗?

20 一般边值问题.

21 具有指定单值群的线性微分方程解的存在性证明.

22 通过自守函数使解析关系单值化.

23 变分法的进一步发展.

在报告的最后,希尔伯特说:"数学的有机的统一是这门科学固有的特点,因为它是一切精确的自然科学知识的基础.为了圆满地实现这个目标,让新世纪给这门科学带来天才的大师和无数热情的信徒吧!"

17.1.5 罗素的悖论和第三次数学危机

从古希腊到现代数学史的研究指出,数学的基础曾受到三次危机的困扰.每一次都是这样,大部分被人们认为确凿无疑的数学受到质疑,并且必须改造.

数学基础的第一次危机起源于公元前五世纪.数学作为一门演绎的科学最早是从泰勒斯开始的,所以这种危机不可能出现得更早.这次危机是由$\sqrt{2}$ 的发现引起的,前面我们已作过详细的介绍.其结果是使数学逐渐走上了演绎科学的道路,为数学采用公理化奠定了基础.

数学基础的第二次危机是十七世纪随着牛顿和莱布尼茨发现微积分而产生的.起初,他们的后继者为这门新数学的威力和应用而陶醉,他们大胆前进,而不管基础是否可靠.随着时间的推移,矛盾和悖论越来越多,数学基础的第二次危机出现了.人们开始认识到,分析大厦原来是建筑在沙滩上的.柯西为解决这次危机迈出了第一步:用精确的极限论代替了模糊的无穷小法.接着,魏尔斯特拉斯及其追随

者们实现了分析的算术化.这时,一般认为,数学基础的第二次危机已经克服,并且,数学的整个结构已被恢复,数学已立于无懈可击的基础之上了.

到19世纪末,康托尔的集合论已经得到数学家们的承认.集合论成功地应用到了其它的数学分支.集合论是数学的基础,由于集合论的使用,数学似乎已经达到了"绝对的严格".但是,正当大家兴高采烈地庆贺数学的绝对严格时,数学王国的大地爆发了另一次强烈的地震.

数学基础的第三次危机是由1897年的突然冲击而出现的.这次危机是由于在康托尔的一般集合论的边缘发现的悖论造成的.因为那么多数学分支都建立在集合论的基础上,所以集合论中悖论的发现自然引起了对数学的整个基本结构的有效性的怀疑.

1897年意大利数学家福蒂(B.Forti)揭示了集合论中的第一个悖论.他的悖论的实质可以用康托尔在两年以后发现的很相似的悖论来描述.康托尔曾证明了:对于任意给定的超限数,总存在一个比它大的超限数,所以不存在最大的超限数.现在考虑这样一个集合,它的元素是所有可能的集合.肯定地,没有一个集合含的元素个数比这个集合的元素的个数多.但是,如果情况果真如此,怎么可能有一个超限数比这个集合的超限数大呢?

福蒂和康托尔的悖论用到集合论的深入结果,但英国数学家罗素(Russell Bertrand 1872—1970)于1902年发现了一个悖论,它除了集合概念本身外不需要别的概念.在描述罗素悖论之前,我们注意下面的事实:一个集合或者是它本身的成员,或者不是它本身的成员.例如,

例1　抽象概念的集合本身是抽象概念,但是,所有人的集合不是一个人.

例2　所有集合的集合本身是一个集合,但是,所有星的集合不是一个星.

我们以 M 表示是它们本身的成员的所有集合的集合，而以 N 表示不是它们本身成员的所有集合的集合. 现在我们问：集合 N 是否是它本身的成员. 如果 N 是他本身的成员，则 N 是 M 的成员，而不是 N 的成员，于是 N 不是它本身的成员. 另一方面如果 N 不是它本身的成员，则 N 是 N 的成员，而不是 M 的成员，于是 N 是它本身的成员. 悖论在于，无论是那一种情况，我们都得到矛盾.

罗素悖论曾以多种形式通俗化. 这些形式中最著名的是罗素在 1919 年给出的，称为理发师悖论. 某村的一个理发师宣称，他给所有不给自己刮脸的人刮脸. 于是出现这样的困境：理发师是否给自己刮脸呢？如果他给自己刮脸，那他就违背了自己的原则；如果他不给自己刮脸，那他就应该为自己刮脸.

罗素的悖论在数学中引起了真正的麻烦. 罗素将他的悖论写信告诉了数理逻辑的先驱弗雷格，而弗雷格正好完成他的关于算术基础的二卷巨著. 弗雷格接到信后，在其著作的末尾伤心地写到："一个科学家遇到的最不愉快的事莫过于，当他的工作完成时，基础崩塌了. 当本书的印刷快要完成时，罗素先生的信就使我陷入这样的境地." 这样就出现了数学史上的第三次数学危机.

第三次数学危机使数学家们意识到，应当建立某种公理系统来对集合论作出必要的规定，以排除"罗素悖论"和其它悖论. 于是数学家们便忙碌起来，不久就出现了好几种公理系统.

康托尔的集合论产生悖论的原因之一是，康托尔的集合论中有"一切集合的集合"的概念. 为了不产生悖论，策梅洛（E. Zermelo，1871—1953）在 1908 年提出一种公理系统，这种公理系统由弗兰克尔（A. A. Fraenkel）在 1921 年加以改进，形成了目前公认的彼此无矛盾的公理系统，简称 ZF 公理系统.

第三次数学危机从整体看来还没有解决到令人满意的程度.

17.1.6 20 世纪初的一场大辩论

ZF 公理系统虽然避开了已知的悖论，但很多数学家对此并不满

意,人们对 ZF 公理系统的逻辑基础还有怀疑,康托尔的连续统假设还没有解决,对选择公理的使用还有怀疑.除去这些问题外还有一个大问题:数学的基础是什么?围绕这些问题,由于哲学观点不同,西方逐渐形成了逻辑主义、直觉主义和形式主义三大派别.他们之间产生了一场大辩论,并把数学推向了一个新阶段.

逻辑主义学派的代表人物是罗素和怀特海(A.N.Whitehead 1861—1947).他们两人合作写了名著《数学原理》三卷,在 1910 年—1913 年出版.他们的基本观点是,**数学即逻辑**.罗素说:"逻辑是数学的青年时代,数学是逻辑的壮年时代".只要不容许"集合的集合"这种逻辑语言出现,悖论就不会发生.

直觉主义学派认为,数学理论的真伪只能用人的直觉去判断.这一派的最早的代表人物是克罗内克.他只承认潜无限的概念,而不承认实无限的概念.他们认为,"全体实数"是不可接受的概念."一切集合的集合"的概念更是不可理解.不承认这些概念的合理性,"悖论"就自然不会出现.

第三派是形式主义,它的代表人物是希尔伯特.他们认为,无论是数学的公理系统或逻辑的公理系统,其中的基本概念都是没有意义的.公理只是一行行符号,无所谓真假,只要能够证明公理系统是相容的,这个公理系统便得到承认,它便代表一种真理.悖论是公理系统不相容的一种表现.

希尔伯特和他的学生们在 1920 到 1930 年间发展了希尔伯特的证明论或元数学,这是建立任何形式系统的相容性的一种方法.从这个思想出发,希尔伯特打算把整个数学都公理化,并证明它的无矛盾性.这一奢望后来被哥德尔打破了.希尔伯特的形式主义计划虽然没有可能全部实现,但他创立的"元数学",已经成为一个重要的数学分支.

迄今为止,这场争论尚未停止.

17.1.7 哥德尔的不完全性定理

1931年在《数学物理月刊》上发表了一篇题为"论《数学原理》和有关系统中的形式不可判定命题"的论文.论文的作者是年仅25岁的奥地利数学家和逻辑学家哥德尔(Kurt Godel 1906—1978),他当时在维也纳大学.论文发表时并没有受到重视,但仅仅过了几年,就受到了专家们的普遍重视,被认为是数学和逻辑的基础方面的划时代文献.

哥德尔的论文指出了公理化过程的局限性,这是人们所始料未及的.他的论文的主要影响有四个方面:

1) 它摧毁了数学的所有重要领域能被完全公理化这一强烈的信念;

2) 它扑灭了沿着希尔伯特曾设想的路线证明数学的内部相容性的全部希望;

3) 它使得人们不得不必须重新评价普遍认可的数学哲学;

4) 它把一个新的、强有力且内容丰富的分析技术引到了基础研究之中.

20世纪早期公理化方法有了蓬勃的发展,人们期望,数学的各个分支都能建立完全的公设集.例如,一般认为,皮亚诺关于自然数系建立的公设集是完全的.但是,哥德尔的论文动摇了这些期望.哥德尔证明了下面的定理:

哥德尔第一定理 对于包含自然数系的任何相容的形式体系F,存在F中的不可判定命题;即存在F中的命题S,使得S和非S都不是在F中可证的.

由此得到,自然数系的任何公设集,如果是相容的就不是完全的.换言之,不管我们能为自然数采用什么样的相容的公设集,总存在关于自然数的命题S,使得S和非S都不能从从这些公设得到证明.这可是令人吃惊的,出乎意料的发现.

数论中有许多著名的猜想,到目前为止,既没有证明也没有推

翻.例如,哥德巴赫猜想,孪生素数是无限的猜想等,都未被证明或推翻.这些猜想是不可判定的命题吗?如果是,那我们就永远不能证明它们.

那么,有没有办法去确定一个命题是不是可判定的呢?也没有!1936年美国逻辑学家车敕(Alonzo Church)证明了下面的定理:

车敕定理　对于包含自然数系的任何相容的形式体系 F,不存在有效的方法,决定 F 中的哪些命题在 F 中是可证的.

这真是使人失望.

希尔伯特一直想证明数学的内部相容性问题,但这也无望,因为哥德尔还证明了下面的定理:

哥德尔第二定理　对于包含自然数系的任何相容的形式体系 F,F 的相容性不能在 F 中被证明.

由此得到,在 F 的不可判定问题中,F 的相容性就是其中的一个.希尔伯特原来的希望是彻底破灭了.

其实,从常识看来这也是自然的.中国有句成语叫做"老王卖瓜,自卖自夸";人们不能听他的自夸,就断定他的瓜是好的.

哥德尔的两条定理指出,任何一个数学分支都做不到完全的公理推演,而且没有一个数学分支能保证自己没有内部矛盾.这真是使数学难堪,数学的真理性又何在呢?

哥德尔的两条定理肯定是所有数学定理中最重要的定理之一.人类对于宇宙和数学地位的认识被迫作出了根本性的改变.数学不再是精确论证的顶峰,不再是真理的化身.数学有它自己的局限性.

撇开这些局限性不谈,数学对人类的贡献仍然是巨大的.它是人类最杰出的智慧结晶,也是人类最富创造性的产物.

答案与提示

第二章

§1

1. 用定理 7.　　2. 反证法. (a) 有等根的条件是 $p^2=4q$;此时 p 必为偶数. (b) 若方程有整数根 x_1,x_2,则 $x_1+x_2=p, x_1x_2=q$. p 为奇数蕴含 x_1,x_2 中一为奇数,一为偶数,其积不可能为奇数.

3. 反证法. 若素数个数是有限的,则可设它们是 $p_1,p_2,\cdots,p_N$. 命 $P=p_1\cdot p_2\cdot\cdots\cdot p_N+1$,易见,任何一个素数 $p_i(i=1,2,\cdots,N)$ 都不是 p 的素因子. 于是,或者 p 是素数,或者 p 有不同于 $p_1,p_2,\cdots,p_N$ 的素因子.

§2

2. 依次从每个无限集中取一个元素,并无限取下去,就可得到一个可数集.

3. 设 M 是一个可数集,N 是 M 的无限子集. 将 M 的全体元素依次排起来,再去掉非 N 的元素,重新编号.

§3

1.　$a+(b-a)e^{i\frac{\pi}{3}}$,或 $a+(b-a)e^{-i\frac{\pi}{3}}$.

2.　$a+(b-a)i, b+(b-a)i$ 或 $a-(b-a)i, b-(b-a)i$ 或 $a+\left(\frac{b-a}{\sqrt{2}}\right)e^{i\frac{\pi}{4}}, a+\left(\frac{b-a}{\sqrt{2}}\right)e^{-i\frac{\pi}{4}}$.

3.　1, $\left[(\sqrt{5}-1)\pm i\sqrt{2\sqrt{5}+10}\right]/4$,

$\left[(-\sqrt{5}-1)\pm i\sqrt{2\sqrt{5}-10}\right]/4$.

4.　由 $z_1+z_2+z_3=0$ 得 $z_3=-(z_1+z_2)$. 取共轭,$\bar{z}_3=-(\bar{z}_1+\bar{z}_2)$. 两式相乘得,

$$z_3\bar{z}_3 = z_1\bar{z}_1 + z_2\bar{z}_2 + z_1\bar{z}_2 + z_2\bar{z}_1,$$

利用$|z_1|^2 = |z_2|^2 = |z_3|^2 = 1$,可得$\bar{z}_1 z_2 + \bar{z}_2 z_1 = -1$.于是,

$$|z_1 - z_2|^2 = (z_1 - z_2)(\bar{z}_1 - \bar{z}_2) = |z_1|^2 + |z_2|^2 - (\bar{z}_1 z_2 + z_1\bar{z}_2) = 3.$$

从而$|z_1 - z_2| = \sqrt{3}$.由此

$$|z_1 - z_2| = |z_2 - z_3| = |z_3 - z_1| = \sqrt{3}.$$

因此,z_1,z_2,z_3是内接于单位圆的正三角形的三个顶点.

第三章

§1

1. $\frac{17}{11} = 1 + \frac{1}{1} + \frac{1}{1} + \frac{1}{5}$,

$3.54 = 3 + \frac{1}{1} + \frac{1}{1} + \frac{1}{5} + \frac{1}{1} + \frac{1}{3}$,

$3.14159 = 3 + \frac{1}{7} + \frac{1}{15} + \frac{1}{1} + \frac{1}{25} + \frac{1}{1} + \frac{1}{7} + \frac{1}{4}$,

2. $3\frac{6}{29}$.

3. $\frac{11}{17} = \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{5}$.

4. $\sqrt{3} = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{1} + \frac{1}{2} + \cdots$.

第四章

§1

3. 证明$3 \mid a(2a^2+7)$.只须对$3 \nmid a$的情况给予证明,故可设$a = 3m \pm 1$.此时

$$2a^2 + 7 = 2(3m \pm 1)^2 + 7 = 2(9m^2 \pm 6m + 1) + 7 = 18m^2 \pm 12m + 9,$$

可为了整除.

4. 令 $a=2m+1$，于是 $(a^2+3)(a^2+7)=16(m^2+m+1)(m^2+m+2)$ 而 m^2+m+1 与 m^2+m+2 是相邻整数，必有一个偶数.

5. 用辗转相除法.

6. 设 p 是素数，α 是自然数.今来证：$\sigma(p^\alpha)<2p^\alpha$.

$$\begin{aligned}2p^\alpha-\sigma(p^\alpha)&=2p^\alpha-(1+p+\cdots+p^\alpha)\\&=p^\alpha-(1+p+\cdots+p^{\alpha-1})\\&=p^\alpha-\frac{p^\alpha-1}{p-1},\end{aligned}$$

因而只需证 $p^\alpha>\dfrac{p^\alpha-1}{p-1}$.因为 $(p-1)\geqslant 1$，所以

$$p^\alpha(p-1)\geqslant p^\alpha>p^\alpha-1$$

从而
$$p^\alpha>\frac{p^\alpha-1}{p-1},$$

即
$$\sigma(p^\alpha)<2p^\alpha.$$

7. 设 $m=p_1^{\alpha_1}\cdots p_n^{\alpha_n}$，则 $\sigma(m^2)=\sigma(p_1^{2\alpha_1})\cdots\sigma(p_n^{2\alpha_n})$ 设 $p_1=2$(当 m 的素因子分解式中不含 2 时，可不考虑它)，则

$$\sigma(2^{2\alpha_1})=1+2+2^2+\cdots+2^{2\alpha_1}=奇数.$$

而 $p_2,p_3,\cdots,p_n$ 皆为奇素数.此时

$$\sigma(p_k^{2\alpha_k})=1+p_k+p_k^2+\cdots+p_k^{2\alpha_k}=奇数\quad(k=2,3,\cdots,n).$$

这样一来，$\sigma(m^2)$ 必为奇数，而 $2m^2$ 是偶数.从而

$$\sigma(m^2)\neq 2m^2.$$

8. 设 p,q 都是奇素数，且 $p>q$.

$$\sigma(pq)=\sigma(p)\sigma(q)=(1+p)(1+q)=1+p+q+pq.$$

只要证
$$1+p+q<pq.$$

因
$$(p-1)(q-1)>2,$$

从而
$$pq>p+q+1.$$

9. $2^{11}-1=23\cdot 89$ 不是素数.由定理6，$2^{10}(2^{11}-1)$ 不是完全数.

第五章

§1

1. 1) $15x+25y=100 \Longleftrightarrow 3x+5y=20$.

特解 $x_0=5, y_0=1$. 通解为

$$x=5-5t,\quad y=1+3t,\quad t=0,\pm1,\pm2,\cdots$$

2) $306x-360y=630 \Longleftrightarrow 17x-20y=35$.

特解: $x_0=-5, y_0=-6$. 通解为

$$x=-5+20t,\quad y=-6+17t,\quad t=0,\pm1,\pm2,\cdots$$

3) $x=t, y=4t-5s, z=21t-24s-200$; $t=0,\pm1,\pm2$; $s=0,\pm1,\pm2,\cdots$.

2. $100=56+44$.

§2

1. 见下表

n \ m	8	9	10
1	(63,16,65)		(99,20,101)
2		(77,36,85)	
3	(55,48,73)		(91,60,109)
4		(65,72,97)	
5	(39,80,89)		
6			

2. 相应的 C 值是 5,13,17,29,37,41,53,61,73,89,97.

3. 证明时要利用等式:

$$x=m^2-n^2,\quad y=2mn,\quad z=m^2+n^2.$$

分两种情况: 1) $3\mid m$ 或 $3\mid n$; 2) $3\nmid mn$

1) $3\mid mn \to 3\mid y$.

2) 令 $m = 3p \pm 1, n = 3q \pm 1$, 我们有

$$\begin{aligned} m^2 - n^2 &= (3p \pm 1)^2 - (3q \pm 1)^2 \\ &= 9p^2 \pm 6p + 1 - (9q^2 \pm 6q + 1) \\ &= 9(p^2 + q^2) \pm 6p \pm 6q. \end{aligned}$$

所以 $3\mid(m^2 - n^2)$, 从而 $3\mid x$.

4. 若 $5\mid m$ 或 $5\mid n$, 则必有 $5\mid y$. 当 $5\nmid m, 5\nmid n$ 时, 可设

$$m = 5t + p, \quad n = 5s + q, \qquad p、q = \pm 1, \pm 2$$

我们有

$$\begin{aligned} m^2 - n^2 &= (5t + p)^2 - (5s + q)^2 \\ &= 25t^2 + 10pt + p^2 - 25s^2 - 10qs - q^2, \end{aligned}$$

当 $p = q = \pm 1, p = q = \pm 2$ 时, $5\mid(m^2 - n^2)$, 即 $5\mid x$,

而 $\qquad m^2 + n^2 = 25t^2 + 10pt + p^2 + 25s^2 + 10qs + q^2$,

当 $p = \pm 1, q = \pm 2$, 或 $p = \pm 2, q = \pm 1$ 时, $5\mid(m^2 + n^2)$, 即 $5\mid z$.

5. (120,22,122). 6. 设有面积为78的毕达哥拉斯三角形, 有一个面积为120的毕达哥拉斯三角形(24,10,26).

7. $1105 = 33^2 + 4^2 = 32^2 + 9^2 = 31^2 + 12^2 = 24^2 + 23^2$, 由此不难求出相应的三角形.

第六章

§2

1. 设 Q_1 由所有形如 $a + b\sqrt{2}$ 的数组成, 其中 a, b 是有理数. 取 $k = \sqrt{2}, \sqrt{k} = \sqrt[4]{2} \notin Q_1$. Q_1 的扩域 Q'中的数只有形式: $p + q\sqrt[4]{2}$, 其中 $p, q \in Q_1$, 即

$$p = a + b\sqrt{2}, \quad q = c + d\sqrt{2}, \qquad a, b, c、d \in Q.$$

容易验证, Q'中的两个数的和、差、积仍在 Q'中, 所需验证的只是两个数的商. 实际上, 只需验证, Q'中的数的倒数仍在 Q'中就够了.

$$\frac{1}{a+b\sqrt[4]{2}} = \frac{a-b\sqrt[4]{2}}{(a+b\sqrt[4]{2})(a-b\sqrt[4]{2})} = \frac{a-b\sqrt[4]{2}}{a^2-b^2\sqrt{2}}$$
$$= \frac{a}{a^2-b^2\sqrt{2}} - \frac{b}{a^2-b^2\sqrt{2}}\sqrt[4]{2},$$

易见,系数

$$\frac{a}{a^2-b^2\sqrt{2}}, \quad \frac{-b}{a^2-b^2\sqrt{2}} \in Q_1.$$

2. $x = \frac{1}{7} - \frac{2}{7}\sqrt{2}, \quad y = \frac{2}{7} + \frac{3}{7}\sqrt{2}.$

第七章

§1

1. 任一整数 a 可表示为

$$a = a_n 10^n + a_{n-1}10^{n-1} + \cdots + a_1 10 + a_0, \quad 0 \leqslant a_i < 10.$$

由 $10 \equiv 1 \pmod 9$ 可推出,$10^i \equiv 1 \pmod 9$,i 是自然数.从而

$$a \equiv a_n + a_{n-1} + \cdots + a_1 + a_0 \pmod 9.$$

2. 同定理 5 的证明.

3. $1535625 = 3^3 \cdot 5^4 \cdot 7 \cdot 13.$

$$1158066 = 2 \cdot 3^2 \cdot 7^2 \cdot 13 \cdot 101$$

4. 对 $n \leqslant 9$,证明提示中的等式.事实上,将左边展开,得等比级数

$$10^n + 10^{n-1} + \cdots + 1 = (10^{n+1} - 1)/9.$$

有了提示中的公式,结果就出来了.

§2

1. (i) $x \equiv 3 \pmod 7$, (ii) $x \equiv 9 \pmod{31}$, (iii) $x \equiv 7 \pmod{21}$, (iv) $(20,30) = 10 \nmid 4$,无解.

2. $x \equiv 2111 \pmod{2310}$.

3. $x \equiv 394 \pmod{1155}$.

§ 3

1. $\varphi(1001)=\varphi(7)\varphi(11)\varphi(13)=720$,
$\varphi(5040)=\varphi(2^4)\varphi(3^2)\varphi(5)\varphi(7)=1152$,
$\varphi(36000)=\varphi(2^5)\varphi(3^2)\varphi(5^3)=9600$.

2. $\varphi(5186)=\varphi(2)\varphi(2593)=2592$,
$\varphi(5187)=\varphi(3)\varphi(7)\varphi(13)\varphi(19)=2592$,
$\varphi(5188)=\varphi(2^2)\varphi(1297)=2592$.

3. 1) 若 n 是奇数,则 $(2,n)=1$.从而

$$\varphi(2n)=\varphi(2)\varphi(n)=\varphi(n).$$

2) 若 n 是偶数,则令 $n=2^\alpha m$, m 为奇数.从而

$$\varphi(2n)=\varphi(2^{\alpha+1})\varphi(m)=(2^{\alpha+1}-2^\alpha)\varphi(m)$$
$$=2(2^\alpha-2^{\alpha-1})\varphi(m).$$

又　$\varphi(n)=\varphi(2^\alpha m)=\varphi(2^\alpha)\varphi(m)=(2^\alpha-2^{\alpha-1})\varphi(m)$

比较两式,得　$\varphi(2n)=2\varphi(n)$.

3) 若 $n=3^\alpha m$,则

$$\varphi(3n)=\varphi(3^{\alpha+1}m)=(3^{\alpha+1}-3^\alpha)\varphi(m)$$
$$=3(3^\alpha-3^{\alpha-1})\varphi(m)=3\varphi(n).$$

4) 若 $(3,n)=1$,则

$$\varphi(3n)=\varphi(3)\varphi(n)=2\varphi(n).$$

4. $\varphi(n)=\varphi(2(2p-1))=\varphi(2)\varphi(2p-1)=\varphi(2p-1)=2p-2$.而

$$\varphi(n+2)=\varphi(2(2p-1)+2)=\varphi(4p)$$
$$=\varphi(4)\varphi(p)=2\varphi(p)=2p-2.$$

5. 星期 4.

6. $5^8\equiv1\pmod 7$, $5^{12}\equiv4\pmod{11}$, $1945^8\equiv1\pmod 7$, $1945^{12}\equiv4\pmod{11}$.

§ 4

2. Sunday

4. 3233 = 53 × 61,复原指数 $j = 253$.

第九章

1. 网络中含二个奇顶点.若不允许重复,则不存在.

2. 可以.

3. 有两个奇顶点,不能.

4. (1) 把每个选手看作一点,不同选手用线连接,表示他们之间有比赛.这样得到一个网络,然后用定理 4.

(2) 由(1),打过 3 盘的选手是偶数,在 225 中除去这一偶数至少剩 1.

5. 这是一个偶网络.

第十章

§1

1. $x_1 \approx 0.454, x_2 \approx -0.227 + 1.468i, x_3 \approx -0.227 - 1.468i$.

2. $x_1 = 1, x_2 = x_3 = -2$.

3. $x_1 \approx 0.229, x_2 \approx 1.385 + 1.564i, x_3 \approx 1.385 - 1.564i$.

4. $x_1 = x_2 = -1, x_3 = \dfrac{-1+\sqrt{13}}{2}, x_4 = \dfrac{-1-\sqrt{13}}{2}$.

§2

1. 1.213.　　2. 2.469.

3. 0.0524.　　4. 0.179.

5. −0.445.　　6. 3.236.

7. 1) 0;2) 4;3) 1.

第十三章

§2

1. 1) $\inf f(x) = 0, \sup f(x) = 25$.

2) $\inf f(x)=0, \sup f(x)=1$.

第十五章

§ 3

1. $y(t)=6.6e^{0.02(t-1988)}$.

2. 200 万.

3. $\frac{4}{5}$

§ 4

1. 2.6×10^{9}.

2. $1622\ln\frac{4}{3}/\ln 2\approx 673$(年).

3. 85.

第十六章

1. 1) $-2dx\wedge dy$,

2) $(x^2\cos x+2yz)dx\wedge dy+x^2zdx\wedge dz+z^2dy\wedge dz$,

3) $-5dx\wedge dy\wedge dz$.

2. 1) $-(\cos y+\sin x)dx\wedge dy$,

2) $(2z+x)dx\wedge dy\wedge dz$.

参考书目录

1　潘承洞，潘承彪．初等数论．北京：北京大学出版社，1992.9

2　胡作玄．从毕达哥拉斯到费尔马．郑州：河南科学技术出版社，1997.1

3　华罗庚．从祖冲之的圆周率谈起．北京：人民教育出版社，1964.2

4　华罗庚．从孙子的《神奇妙算》谈起．北京：人民教育出版社，1964.3

5　姜伯驹．一笔画．北京：人民教育出版社，1964

6　M.克莱因．李宏魁译．数学：确定性的丧失．长沙：湖南科学技术出版社，1997.6

7　西蒙·辛格．薛密译．费马大定理．上海：上海译文出版社，1998.2

8　Robert Devaney Chaos, Fractal and Dynamics. Computer Experiments in Mathematics. Addison-Wesley, 1990

9　Martin Braun. Differential Equations and Their Applications. Fourth Edition. New York: Springer-Verlag, 1993